ASM HANDBOOK®

VOLUME

18

Friction, Lubrication, and Wear Technology

ASM INTERNATIONAL®

The Materials Information Company

Publication Information and Contributors

Friction, Lubrication, and Wear Technology was published in 1992 as Volume 18 of the *ASM Handbook*. The Volume was prepared under the direction of the ASM International Handbook Committee.

Volume Chair

The Volume Chairman was Peter J. Blau, Metals and Ceramics Division, Oak Ridge National Laboratory.

Authors

- **Arnold E. Anderson** Consultant
- **Walter K. Arnold** Fraunhofer Institute
- **Betzalel Avitzur** Metalforming Inc.
- **Stephen C. Bayne** University of North Carolina
- **Charles C. Blatchley** Spire Corporation
- **Peter J. Blau** Oak Ridge National Laboratory
- **Raymond H. Boehringer** DuBois Chemical Inc.
- **Royce N. Brown** Dow Chemical U.S.A.
- **Kenneth G. Budinski** Eastman Kodak Company
- **R.F. Bunshah** University of California, Los Angeles
- **Ralph A. Burton** Burton Technologies Inc.
- **Herbert S. Cheng** Northwestern University
- **Stanley Chinowsky** Pure Carbon Company
- **Y.-W. Chung** Northwestern University
- **Robert D. Compton** Noran Instruments Inc.
- **J.M. Conway-Jones** Glacier Vandervell Inc.
- **Khershed P. Cooper** Naval Research Laboratory
- **Richard S. Cowan** Georgia Institute of Technology
- **Paul Crook** Haynes International Inc.
- **Carl E. Cross** Martin Marietta
- **H. Czichos** Bundesanstalt für Materialforschung und -Prüfung (BAM)
- **Raymond J. Dalley** Predict Technologies
- **Steven Danyluk** University of Illinois at Chicago
- **Mark Davidson** University of Florida
- **Joseph R. Davis** Davis & Associates
- **Duncan Dowson** University of Leeds
- **James F. Dray** Mechanical Technology Inc.
- **David M. Eissenberg** Oak Ridge National Laboratory
- **Peter A. Engel** State University of New York at Binghamton
- **Robert Errichello** Geartech
- **Terry S. Eyre** Eyre Associates
- **Howard N. Farmer** Haynes International Inc.
- **Richard S. Fein** Fein Associates
- **George R. Fenske** Argonne National Laboratory
- **Paul D. Fleischauer** Aerospace Corporation
- **Dudley D. Fuller** Columbia University
- **William A. Glaeser** Battelle Memorial Institute
- **Douglas A. Granger** Aluminum Company of America

- **Austin L. Grogan, Jr.** University of Central Florida
- **Inge L.H. Hansson** Alcan International Ltd.
- **Carolyn M. Hansson** Queen's University
- **Tedric A. Harris** Pennsylvania State University
- **Howard D. Haynes** Oak Ridge National Laboratory
- **Per Hedenqvist** Uppsala University
- **Frank J. Heymann** Consultant
- **Michael R. Hilton** Aerospace Corporation
- **Franz Hoffmann** Stiftung Institut für Werkstofftechnik
- **Sture Hogmark** Uppsala University
- **Roger G. Horn** National Institute of Standards and Technology
- **C.R. Houska** Virginia Polytechnic Institute
- **Lewis K. Ives** National Institute of Standards and Technology
- **Staffan Jacobsson** Uppsala University
- **William R. Kelley** Borg-Warner Automotive
- **L. Alden Kendall** University of Minnesota, Duluth
- **Francis E. Kennedy, Jr.** Dartmouth College
- **George R. Kingsbury** Glacier Vandervell Inc.
- **Thomas H. Kosel** University of Notre Dame
- **Burton A. Kushner** Metco/Perkin-Elmer
- **Frank M. Kustas** Martin Marietta Aerospace
- **Joseph T. Laemmle** Aluminum Company of America
- **Jorn Larsen-Basse** National Science Foundation
- **Soo-Wohn Lee** University of Illinois at Chicago
- **A.V. Levy** Lawrence Berkeley Laboratory
- **Y. Liu** University of Wisconsin-Milwaukee
- **Frances E. Lockwood** Pennzoil Products Company
- **Kenneth C. Ludema** University of Michigan
- **Brent W. Madsen** U.S. Bureau of Mines
- **John H. Magee** Carpenter Technology Corporation
- **James L. Maloney III** Latrobe Steel
- **William D. Marscher** Dresser Industries
- **Hugh R. Martin** University of Waterloo
- **P. Mayr** Stiftung Institut für Werkstofftechnik
- **John E. Miller** White Rock Engineering Inc.
- **Mohan S. Misra** Martin Marietta Aerospace
- **Charles A. Moyer** Timken Company
- **U. Netzelmann** Fraunhofer Institute
- **Edward R. Novinski** Metco/Perkin-Elmer
- **David L. Olson** Colorado School of Mines
- **Michael Olsson** Uppsala University
- **S. Pangraz** Fraunhofer Institute
- **Ron Pike** Glacier Vandervell Inc.
- **Padmanabha S. Pillai** Goodyear Tire & Rubber Company
- **Hubert M. Pollock** Lancaster University
- **John M. Powers** University of Texas
- **Terence F.J. Quinn** United States International University
- **S. Ray** University of Wisconsin-Milwaukee
- **Stephen L. Rice** University of Central Florida
- **Syed Q.A. Rizvi** Lubrizol Corporation
- **Pradeep Rohatgi** University of Wisconsin-Milwaukee
- **A.W. Ruff** National Institute of Standards and Technology
- **John Rumierz** SKF USA Inc.
- **Leonard E. Samuels** Samuels Consultants

- **Jerry D. Schell** General Electric Aircraft Engines
- **Monica A. Schmidt** Martin Marietta Energy Systems Inc.
- **Henry J. Scussel** GTE Valenite
- **S.L. Semiatin** Wright Laboratory
- **Barrie S. Shabel** Aluminum Company of America
- **Keith Sheppard** Stevens Institute of Technology
- **Rajiv Shivpuri** Ohio State University
- **Harold E. Sliney** NASA Lewis Research Center
- **J.F. Song** National Institute of Standards and Technology
- **T.S. Sriram** Northwestern University
- **Charles A. Stickels** Environmental Research Institute of Michigan
- **E.M. Tatarzycki** Aircraft Braking Systems Corporation
- **Kevin P. Taylor** General Electric Aircraft Engines
- **William G. Truckner** Aluminum Company of America
- **Joseph H. Tylczak** U.S. Bureau of Mines
- **Olof Vingsbo** Uppsala University
- **T.V. Vorburger** National Institute of Standards and Technology
- **Robert B. Waterhouse** University of Nottingham
- **R.T. Webb** Aircraft Braking Systems Corporation
- **Rolf Weil** Stevens Institute of Technology
- **Eric P. Whitenton** National Institute of Standards and Technology
- **Ward O. Winer** Georgia Institute of Technology

Reviewers and Contributors

- **Taylan Altan** Ohio State University
- **Doug Asbury** Cree Research
- **Shyam Bahadur** Iowa State University
- **Randall F. Barron** Louisiana Tech University
- **Raymond Bayer** Consultant
- **Abdel E. Bayoumi** Washington State University
- **Horst Becker** Sintermet Corporation
- **Charles Bellanca** Dayton Power and Light
- **Robert K. Betts** Cincinnati Thermal Spray Inc.
- **Peter J. Blau** Oak Ridge National Laboratory
- **Rodney R. Boyer** Boeing Commercial Airplane Group
- **Robert W. Bruce** General Electric Aircraft Engines
- **Gerald Bruck** Westinghouse STC
- **Michael Bryant** University of Texas
- **R.A. Buchanan** University of Tennessee, Knoxville
- **Kenneth G. Budinski** Eastman Kodak Company
- **Harold I. Burrier, Jr.** Timken Company
- **Donald C. Carmichael** Battelle Memorial Institute
- **J.A. Carpenter, Jr.** National Institute of Standards and Technology
- **A.G. Causa** Goodyear Tire & Rubber Company
- **Y.P. Chiu** Torrington Company
- **Ronald Christy** Tribo Coating
- **Richard S. Cowan** Georgia Institute of Technology
- **W.J. Crecelius** General Electric
- **G.R. Crook** Aluminum Company of America
- **Bob Dawson** Deloro Stellite Inc.
- **Arnold O. DeHart** Bearing Systems Technology
- **Christopher DellaCorte** NASA Lewis Research Center
- **Paolo DeTassis** Clevite SpA

- **John Deuber** Degussa Corporation
- **Mitchell R. Dorfman** Metco/Perkin-Elmer
- **Keith Dufrane** Battelle Memorial Institute
- **Lawrence D. Dyer** Dyer Consultants
- **Norman S. Eiss, Jr.** Virginia Polytechnic Institute and State University
- **Wayne L. Elban** Loyola College
- **T.N. Farris** Purdue University
- **Neal Fechter** National Electric Carbon Corporation
- **Andrew Fee** Wilson Instruments
- **Richard S. Fein** Fein Associates
- **Gregory A. Fett** Dana Corporation
- **Traugott Fischer** Stevens Institute of Technology
- **Donald G. Flom** Flom Consulting
- **Anna C. Fraker** National Institute of Standards and Technology
- **Steven G. Fritz** Southwest Research Institute
- **Raymond P. Funk** Cato Oil & Grease Company
- **Michelle M. Gauthier** Raytheon Company
- **Louis T. Germinario** Eastman Chemical Company
- **S.K. Ghosh** Eastman Kodak Company
- **W.A. Glaeser** Battelle Memorial Institute
- **E.W. Glossbrenner** Litton Poly-Scientific
- **Allan E. Goldman** U.S. Graphite Inc.
- **Steven Granick** University of Illinois
- **Robert E. Green, Jr.** Johns Hopkins University
- **Walter P. Groff** Southwest Research Institute
- **John J. Groth** FMC Corporation
- **Raymond A. Guyer, Jr.** Rolling Bearing Institute Ltd.
- **Tom Heberling** Armco Inc. Research Laboratories
- **Frank Heymann** Consultant
- **Robert Hochman** Georgia Institute of Technology
- **James C. Holzwarth** General Motors Research Laboratories (Retired)
- **Hyun-Soo Hong** Lubrizol Corporation
- **James Hudson** A-C Compressor Corporation
- **Allan B. Hughes** Actis Inc.
- **S. Ibarra** Amoco Corporation Research
- **J. Ernesto Indacochea** University of Illinois at Chicago
- **Said Jahanmir** National Institute of Standards and Technology
- **Bob Jaklevic** Ford Motor Company
- **Kishore Kar** Dow Chemical Company
- **Igor J. Karassik** Dresser Pump Division, Dresser Industries
- **Francis E. Kennedy, Jr.** Dartmouth College
- **M.K. Keshavan** Smith International
- **L.L. Kesmodel** Indiana University
- **Paul Y. Kim** National Research Council
- **George Krauss** Colorado School of Mines
- **Jorn Larsen-Basse** National Science Foundation
- **P.W. Lee** Timken Company
- **Minyoung Lee** General Electric Company
- **Herman R. Leep** University of Louisville
- **Kenneth Liebler**
- **Richard Lindeke** University of Minnesota
- **Walter E. Littmann** Failure Analysis Associates Inc.
- **Stephen Liu** Colorado School of Mines
- **Frances E. Lockwood** Pennzoil Products Company

- **Robert A. Lord** Dresser-Rand Company
- **William Lucke** Cincinnati Milacron
- **Kenneth C. Ludema** University of Michigan
- **William L. Mankins** Inco Alloys International Inc.
- **Jacques Masounave** E.T.S. Université du Québec
- **I.D. Massey** Glacier Vandervell Ltd.
- **P.M. McGuiggan** 3M Company
- **Paul Mehta** General Electric Aircraft Engines
- **John E. Miller** White Rock Engineering Inc.
- **John C. Mitchem** Oregon Health Sciences University
- **K. Miyoshi** NASA Lewis Research Center
- **P.A. Molian** Iowa State University
- **Dave Neff** Metaullics Systems
- **Welville B. Nowak** Northeastern University
- **Han Nyo** BP Chemicals (Hitco) Inc.
- **Warren Oliver** Oak Ridge National Laboratory
- **David L. Olson** Colorado School of Mines
- **Daniel W. Parker** General Plasma
- **Konrad Parker** Consultant
- **Sanjay Patel** AT&T Bell Laboratories
- **Burton R. Payne, Jr.** Payne Chemical Corporation
- **Marshall B. Peterson** Wear Sciences Corporation
- **William W. Poole** United Technologies Corporation
- **Marion L. Pottinger** Smithers Scientific Services Inc.
- **K. Prewo** United Technologies Research Center
- **C. Pulford** Goodyear Tire & Rubber Company
- **J. Raja** University of North Carolina at Charlotte
- **Seong K. Rhee** Allied-Signal Friction Materials
- **Stephen L. Rice** University of Central Florida
- **David A. Rigney** Ohio State University
- **Gary Rimlinger** Aircraft Braking Systems Corporation
- **Michael L. Rizzone** Consulting Mechanical Engineer
- **Elwin L. Rooy** Consultant
- **Jules Routbort** Argonne National Laboratory
- **A.W. Ruff** National Institute of Standards and Technology
- **Nannaji Saka** Massachusetts Institute of Technology
- **Ronald O. Scattergood** North Carolina State University
- **J.A. Schey** University of Waterloo
- **George F. Schmitt, Jr.**
- **William Schumacher** Armco Research & Technology
- **Christopher G. Scott** Lubrizol Corporation
- **Wilbur Shapiro** Mechanical Technology Inc.
- **Hal Shaub** Exxon Chemical Company
- **M.C. Shaw** Arizona State University
- **Lewis B. Sibley** Tribology Systems Inc.
- **Fred A. Smidt** Naval Research Laboratory
- **Darrell W. Smith** Michigan Technological University
- **Talivaldis Spalvins** NASA Lewis Research Center
- **Cullie J. Sparks, Jr.** Oak Ridge National Laboratory
- **Donald R. Spriggs** Chem-tronics Aviation Repair
- **Karl J. Springer** Southwest Research Institute
- **William D. Sproul** BIRL Northwestern University
- **D.S. Stone** University of Wisconsin
- **W. Sutton** United Technologies Research Center

- **Shoji Suzuki** Asahi Glass America Inc.
- **Paul A. Swanson** Deere & Company
- **Roderic V. Sweet** MRC Bearing Services
- **A.R. Thangaraj** Michigan Technological University
- **Frank Toye** Leco Corporation
- **Ronald L. Trauger**
- **George Vander Voort** Carpenter Technology Corporation
- **William von Kampen** General Motors Truck & Bus
- **Roy Waldheger** Carbon Technology Inc.
- **Malcolm J. Werner** Bently Nevada
- **Grady S. White** National Institute of Standards and Technology
- **Eric P. Whitenton** National Institute of Standards and Technology
- **Douglas D. Wilson** Friction Products Company
- **Ward O. Winer** Georgia Institute of Technology
- **Jerry O. Wolfe** Timken Company
- **William A. Yahraus** Failure Analysis Associates Inc.
- **William B. Young** Dana Corporation
- **Charles S. Yust** Oak Ridge National Laboratory
- **G. Zajac** Amoco Research Center
- **Dong Zhu** Alcoa Technical Center

Foreword

The publication of this Volume marks the first time that the *ASM Handbook* has dealt with friction, lubrication, and wear technology as a separate subject. However, the tribological behavior of materials and components has been of fundamental importance to ASM members throughout the history of the Society. ASM International traces its origins back to 1913 with the formation of the Steel Treaters Club in Detroit. This group joined with the American Steel Treaters Society to form the American Society for Steel Treating in 1920. In the early history of the Society as an organization devoted primarily to heat treating, one of the key interests of its membership was improving the wear properties of steel.

In 1933 the organization changed its name to the American Society for Metals, completing its transformation to an organization that served the interests of the entire metals industry. This change led the Society into many other areas--such as metalworking, surface finishing, and failure analysis--where friction, lubrication, and wear are key concerns. In 1987 the technical scope of the Society was further broadened to include the processing, properties, and applications of all engineering/structural materials, and thus ASM International was born. This Handbook reflects the wide focus of the Society by addressing the tribological behavior of a broad range of materials.

The comprehensive coverage provided by this Volume could not have been achieved without the planning and coordination of Volume Chairman Peter J. Blau. He has been tireless in his efforts to make this Handbook the most useful tool possible. Thanks are also due to the Section Chairmen, to the members of the ASM Handbook Committee, and to the ASM editorial staff. We are especially grateful to the over 250 authors and reviewers who so generously donated their time and expertise to make this Handbook an outstanding source of information.

- William P. Koster
  President
  ASM International
- Edward L. Langer
  Managing Director
  ASM International

Preface

Friction, lubrication, and wear (FL&W) technology impacts many aspects of daily life, from the wear of one's teeth to the design of intricate, high-speed bearings for the space shuttle. Nearly everyone encounters a FL&W problem from time to time. Sometimes the solution to the problem is simple and obvious--disassembling, cleaning, and relubricating a door

hinge, for example. Sometimes, however, the problem itself is difficult to define, the contact conditions in the system difficult to characterize, and the solution elusive. Approaches to problem-solving in the multidisciplinary field of tribology (that is, the science and technology of FL&W) often present a wide range of options and can include such diverse fields as mechanical design, lubrication, contact mechanics, fluid dynamics, surface chemistry, solid-state physics, and materials science and engineering. Practical experience is a very important resource for solving many types of FL&W problems, often replacing the application of rigorous tribology theory or engineering equations. Selecting "the right tool for the right job" was an inherent principle in planning the contents of this Volume.

It is unrealistic to expect that specific answers to all conceivable FL&W problems will be found herein. Rather, this Handbook has been designed as a resource for basic concepts, methods of laboratory testing and analysis, materials selection, and field diagnosis of tribology problems. As Volume Chairman, I asked the Handbook contributors to keep in mind the question: "What information would I like to have on my desk to help me with friction, lubrication, or wear problems?" More than 100 specialized experts have risen to this challenge, and a wealth of useful information resides in this book.

The sections on solid friction, lubricants and lubrication, and wear and surface damage contain basic, tutorial information that helps introduce the materials-oriented professional to established concepts in tribology. The Handbook is also intended for use by individuals with a background in mechanics or lubricant chemistry and little knowledge of materials. For example, some readers may not be familiar with the measurement and units of viscosity or the regimes of lubrication, and others may not know the difference between brass and bronze. The "Glossary of Terms" helps to clarify the use of terminology and jargon in this multidisciplinary area. The discerning reader will find the language of FL&W technology to be somewhat imprecise; consequently, careful attention to context is advised when reading the different articles in the Volume.

The articles devoted to various laboratory techniques for conducting FL&W analyses offers a choice of tools to the reader for measuring wear accurately, using these measurements to compute wear rates, understanding and interpreting the results of surface imaging techniques, and designing experiments such that the important test variables have been isolated and controlled. Because many tribosystems contain a host of thermal, mechanical, materials, and chemical influences, structured approaches to analyzing complex tribosystems have also been provided.

The articles devoted to specific friction- or wear-critical components are intended to exemplify design and materials selection strategies. A number of typical tribological components or classes of components are described, but it was obviously impossible to include all the types of moving mechanical assemblies that may experience FL&W problems. Enough diversity is provided, however, to give the reader a solid basis for attacking other types of problems. The earlier sections dealing with the basic principles of FL&W science and technology should also be useful in this regard.

Later sections of the Handbook address specific types of materials and how they react in friction and wear situations. Irons, alloy steels, babbitts, and copper alloys (brasses and bronzes) probably account for the major tonnage of tribological materials in use today, but there are technologically important situations where these workhorse materials may not be appropriate. Readers with tribomaterials problems may find the sections on other materials choices, such as carbon-graphites, ceramics, polymers, and intermetallic compounds, helpful in providing alternate materials-based solutions. In addition, the section on surface treatments and modifications should be valuable for attacking specialized friction and wear problems. Again, the point is to find the right material for the right job.

This Volume marks the first time that ASM International has compiled a handbook of FL&W technology. The tribology research and development community is quite small compared with other disciplines, and the experts who agreed to author articles for this Volume are extremely busy people. I am delighted that such an outstanding group of authors rallied to the cause, one that ASM and the entire tribology community can take pride in. I wish to thank all the contributors heartily for their much-appreciated dedication to this complex and important project in applied materials technology.

- Peter J. Blau, Volume Chairman
  Metals and Ceramics Division
  Oak Ridge National Laboratory

General Information

Officers and Trustees of ASM International (1991-1992)

- **William P. Koster** President and Trustee Metcut Research Associates Inc.
- **Edward H. Kottcamp, Jr.** Vice President and Trustee SPS Technologies
- **Stephen M. Copley** Immediate Past President and Trustee Illinois Institute of Technology
- **Edward L. Langer** Secretary and Managing Director ASM International
- **Leo G. Thompson** Treasurer Lindberg Corporation
- **Trustees**
- **William H. Erickson** Canada Centre for Minerals & Energy Technology
- **Norman A. Gjostein** Ford Motor Company
- **Nicholas C. Jessen, Jr.** Martin Marietta Energy Systems, Inc
- **E. George Kendall** Northrop Aircraft
- **George Krauss** Colorado School of Mines
- **Kenneth F. Packer** Packer Engineering, Inc.
- **Hans Portisch** VDM Technologies Corporation
- **Lyle H. Schwartz** National Institute of Standards and Technology
- **John G. Simon** General Motors Corporation

Members of the ASM Handbook Committee (1991-1992)

- **David LeRoy Olson**(Chairman 1990-; Member 1982-1988; 1989-) Colorado School of Mines
- **Ted Anderson** (1991-) Texas A&M University
- **Roger J. Austin** (1984-) Hydro-Lift
- **Robert J. Barnhurst** (1988-) Noranda Technology Centre
- **John F. Breedis** (1989-) Olin Corporation
- **Stephen J. Burden** (1989-) GTE Valenite
- **Craig V. Darragh** (1989-) The Timken Company
- **Russell J. Diefendorf** (1990-) Clemson University
- **Aicha Elshabini-Riad** (1990-) Virginia Polytechnic & State University
- **Michelle M. Gauthier** (1990-) Raytheon Company
- **Toni Grobstein** (1990-) NASA Lewis Research Center
- **Susan Housh** (1990-) Dow Chemical U.S.A.
- **Dennis D. Huffman** (1982-) The Timken Company
- **S. Jim Ibarra** (1991-) Amoco Research Center
- **J. Ernesto Indacochea** (1987-) University of Illinois at Chicago
- **Peter W. Lee** (1990-) The Timken Company
- **William L. Mankins** (1989-) Inco Alloys International, Inc.
- **David V. Neff** (1986-) Metaullics Systems
- **Richard E. Robertson** (1990-) University of Michigan
- **Elwin L. Rooy** (1989-) Consultant
- **Jeremy C. St. Pierre** (1990-) Hayes Heat Treating Corporation
- **Ephraim Suhir** (1990-) AT&T Bell Laboratories
- **Kenneth Tator** (1991-) KTA-Tator, Inc.
- **William B. Young** (1991-) Dana Corporation

Previous Chairmen of the ASM Handbook Committee

- **R.S. Archer** (1940-1942) (Member, 1937-1942)
- **L.B. Case** (1931-1933) (Member, 1927-1933)
- **T.D. Cooper** (1984-1986) (Member, 1981-1986)
- **E.O. Dixon** (1952-1954) (Member, 1947-1955)
- **R.L. Dowdell** (1938-1939) (Member, 1935-1939)
- **J.P. Gill** (1937) (Member, 1934-1937)
- **J.D. Graham** (1966-1968) (Member, 1961-1970)
- **J.F. Harper** (1923-1926) (Member, 1923-1926)

- **C.H. Herty, Jr.** (1934-1936) (Member, 1930-1936)
- **D.D. Huffman** (1986-1990) (Member, 1990-)
- **J.B. Johnson** (1948-1951) (Member, 1944-1951)
- **L.J. Korb** (1983) (Member, 1978-1983)
- **R.W.E. Leiter** (1962-1963) (Member, 1955-1958, 1960-1964)
- **G.V. Luerssen** (1943-1947) (Member, 1942-1947)
- **G.N. Maniar** (1979-1980) (Member, 1974-1980)
- **J.L. McCall** (1982) (Member, 1977-1982)
- **W.J. Merten** (1927-1930) (Member, 1923-1933))
- **N.E. Promisel** (1955-1961) (Member, 1954-1963)
- **G.J. Shubat** (1973-1975) (Member, 1966-1975)
- **W.A. Stadtler** (1969-1972) (Member, 1962-1972)
- **R. Ward** (1976-1978) (Member, 1972-1978)
- **M.G.H. Wells** (1981) (Member, 1976-1981)
- **D.J. Wright** (1964-1965) (Member, 1959-1967)

Staff

ASM International staff who contributed to the development of the Volume included Scott D. Henry, Editor, ASM Handbooks; Grace M. Davidson, Production Project Manager; Theodore B. Zorc, Technical Editor; Dawn Levicki, Editorial Assistant; Robert C. Uhl, Director of Reference Publications. Editorial assistance was provided by Joseph R. Davis, Kelly Ferjutz, Heather Lampman, Kathleen M. Mills, Nikki D. Wheaton, and Mara S. Woods.

Conversion to Electronic Files

ASM Handbook, Volume 18, *Friction, Lubrication, and Wear Technology* was converted to electronic files in 1997. The conversion was based on the Second Printing (March 1995). No substantive changes were made to the content of the Volume, but some minor corrections and clarifications were made as needed.

ASM International staff who contributed to the conversion of the Volume included Sally Fahrenholz-Mann, Bonnie Sanders, Scott Henry, Grace Davidson, Randall Boring, Robert Braddock, Kathleen Dragolich, and Audra Scott. The electronic version was prepared under the direction of William W. Scott, Jr., Technical Director, and Michael J. DeHaemer, Managing Director.

Copyright Information (for Print Volume)

Copyright © 1992 by ASM International

All Rights Reserved.

ASM Handbook is a collective effort involving thousands of technical specialists. It brings together in one book a wealth of information from world-wide sources to help scientists, engineers, and technicians solve current and long-range problems.

Great care is taken in the compilation and production of this Volume, but it should be made clear that no warranties, express or implied, are given in connection with the accuracy or completeness of this publication, and no responsibility can be taken for any claims that may arise.

Nothing contained in the *ASM Handbook* shall be construed as a grant of any right of manufacture, sale, use, or reproduction, in connection with any method, process, apparatus, product, composition, or system, whether or not covered by letters patent, copyright, or trademark, and nothing contained in the *ASM Handbook* shall be construed as a defense against any alleged infringement of letters patent, copyright, or trademark, or as a defense against liability for such infringement.

Comments, criticisms, and suggestions are invited, and should be forwarded to ASM International.

Library of Congress Cataloging-in-Publication Data (for Print Volume)

ASM International

ASM Handbook.

Title proper has changed with v.4: ASM Handbook.

Vol. 18: Prepared under the direction of the ASM International Handbook Committee. Includes bibliographies and indexes. Contents: v. 18. Friction, lubrication, and wear technology

1. Metals--Handbooks, manuals, etc. I. ASM International. Handbook Committee. II. Title: ASM Handbook.

TA459.M43 1990 620.1'6 90-115

ISBN 0-87170-377-7 (v.1)

SAN 204-7586

ISBN 0-87170-380-7

Printed in the United States of America

Introduction to Friction

Jorn Larsen-Basse, National Science Foundation

FRICTION is the resistance to movement of one body over body. The word comes to us from the Latin verb *fricare,* which means to rub. The bodies in question may be a gas and a solid (aerodynamic friction), or a liquid and a solid (liquid friction); or the friction may be due to internal energy dissipation processes within one body (internal friction). In this article, the discussion will be limited to the effects of solid friction.

Two of the most significant inventions of early man are friction-related: He learned to use frictional heating to start his cooking fires, and he discovered that rolling friction is much less than sliding friction (that is, it is easier to move heavy objects if are on rollers than it is to drag them along). This second discovery would eventually lead to the invention of the wheel.

Friction plays an important role in a significant number of our daily activities and in most industrial processes. It aids in starting the motion of a body, changing its direction, and subsequently stopping it. Without friction, we could not readily move about, grip objects, light a match, or perform a multitude of other common daily tasks. Without friction, most threaded joints would not hold, rolling mills could not operate, and friction welding would obviously not exist. Without friction, we would hear neither the song of the violin nor the squeal of the brake.

In moving machinery, friction is responsible for dissipation and loss of much energy. It has been estimated, for example, that 10% of oil consumption in the United States is used simply to overcome friction. The energy lost to friction is an energy input that must continually be provided in order to maintain the sliding motion. This energy is dissipated in the system, primarily as heat--which may have to be removed by cooling to avoid damage and may limit the conditions under which the machinery can be operated. Some of the energy is dissipated in various deformation processes, which result in wear of the sliding surfaces and their eventual degradation to the point where replacement of whole components becomes necessary. Wear of sliding surfaces adds another, very large component to the economic importance of friction, because without sliding friction these surfaces would not wear.

The fundamental experimental laws that govern friction of solid bodies are quite simple. They are usually named for Coulomb, who formulated them in 1875 (much of his work was built on earlier work by Leonardo da Vinci and Amontons). The laws can be stated in very general terms:

- Static friction my be greater than kinetic (or dynamic) friction
- Friction is independent of sliding velocity
- Friction force is proportional to applied load
- Friction force is independent of contact area

It must be emphasized that these "laws" are very general in nature and that, while they are applicable in many instances, there are also numerous conditions under which they break down.

Friction is commonly represented by the friction coefficient, for which the symbols μ or *f* generally are used. The friction coefficient is the ratio between the friction force, *F,* and the load, *N:*

$$\mu = \frac{F}{N} \qquad \textbf{(Eq 1)}$$

The friction coefficient typically ranges from 0.03 for a very well lubricated bearing, to 0.5 to 0.7 for dry sliding, and even $\geq$5 for clean metal surfaces in a vacuum. A μ-value of 0.2 to 0.3 allows for comfortable walking; however, walking on ice is very difficult because the μ-value for the ice/shoe pair may be <0.05, and a slippery floor may have a μ-value of 0.15. Nature has provided highly efficient lubrication to another component of walking, the knee joint, which has a μ-

value of 0.02. A. representative list of typical friction coefficients is given in the article "Appendix: Static and Kinetic Friction Coefficients for Selected Materials" in this Volume.

A body of weight *W* on a flat surface will begin to move when the surface is tilted to a certain angle (the friction angle, θ) (Fig. 1). The static friction coefficient is given by

$$\mu_s = \tan\theta \qquad \textbf{(Eq 2)}$$

This represents a simple way to measure μ_s, but force measurements are some generally used to measure both the static and the dynamic, or kinetic, coefficients of friction. The results obtained from these measurements do, however, depend on the nature and cleanliness of the surfaces and also to some extent on the various characteristics of the measuring system. This dependence underscores the basic fact that the friction coefficient is not a unique, clearly defined materials property, as may become evident from the following brief discussion of the basic mechanisms of friction

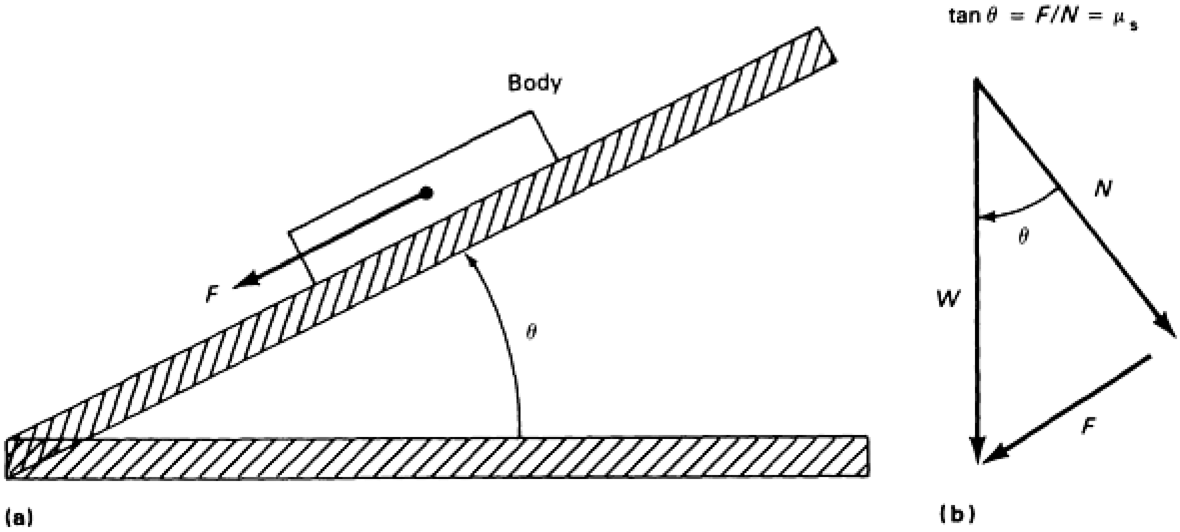


Fig. 1 Inclined plane used to determine coefficient of static friction, μ_s. (a) Tilting flat surface through smallest angle, θ, needed to initiate movement of the body down the plane. (b) Relation of the friction angle to the principal applied forces

Surfaces are not completely flat at the microscopic level. At high magnification, even the best polished surface will show ridges and valleys, asperities, and depressions. When two surfaces are brought together, they touch intimately only at the tips of a few asperities. At these points, the contact pressure may be close to the hardness of the softer material; plastic deformation takes place on a very local scale, and cold welding may form strongly bonded junctions between the two materials. When sliding begins, these junctions have to be broken by the friction force, and this provides the adhesive component of the friction. Some asperities may plow across the surface of the mating material, and the resulting plastic deformation or elastic hysteresis contribute to the friction force; additional contributions may be due to wear by debris particles that become trapped between the sliding surfaces.

Because so many mechanisms are involved in generating the friction force, it is clear that friction is not a unique materials property, but instead depends to some extent on the measuring conditions, on the surface roughness, on the presence or absence of oxides or adsorbed films, and so on. In spite of this complexity, the values of μ obtained by different methods and by different laboratories tend to fall into ranges that are representative of the material pair in question under reasonably similar conditions. That is, values obtained by different laboratories tend to fall within ~20 to 30% of each other if the testing conditions are generally similar. It is important, however, to understand that the values of μ listed in this Handbook are intended only to provide rough guidelines and that more exact values, if needed, must be obtained from direct measurements on the system in question under its typical operating conditions. Detailed information on friction measurement techniques is available in the article "Laboratory Testing Methods for Solid Friction" in this Volume.

The deformation at asperities and junctions is extremely localized, and very high temperatures may therefore be generated over very short periods of time. At these local hod spots, rapid oxidation, plastic flow, or interdiffusion can take place, and these all affect the wear process. In some cases, sparks may even form. The temperatures obtained depend on how fast heat is generated (that is, on the operating conditions of load and velocity) and on how fast heat is removed (that is, on the thermal properties of the sliding surfaces). These temperatures can be calculated with some degree of certainty, as shown in the article "Frictional Heating Calculations" in this Volume.

Friction oscillations may develop when the static coefficient of friction is greater than the kinetic, as is the case for many unlubricated systems. The resulting motion is often called "stick-slip." The two surfaces stick together until the elastic energy of the system has built up to the point where a sudden forward slip takes place. The resulting oscillations may produce equipment vibrations, surface damage, and noise.

Some of the areas of current technological interest and research related to friction include:

- *Friction Measurement:* More accurate ways to measure μ and to predict its value for given conditions without having to test the actual system
- *Friction Sensing:* Use of the various signals that are generated by friction for real-time feedback control of robots, manufacturing processes, lubrication systems, and so on
- *Materials:* Materials and coatings with low friction for operation at elevated temperatures where normal lubricants break down; and materials and coatings with constant, predictable, and sustainable values of μ

Selected References

- F.P. Bowden and D. Tabor, *Friction and Lubrication,* 2nd ed., Methuen, 1964
- F.P. Bowden and D. Tabor, *Friction. An Introduction to Tribology,* Robert Krieger Publishing, 1982
- D. Dowson, *History of Tribology,* Oxford University, Oxford, 1979
- E. Rabinowicz, *Friction and Wear of Materials,* Wiley, 1965
- E. Rabinowicz, Friction, *McGraw-Hill Concise Encyclopedia of Science and Technology,* McGraw-Hill, 1984
- W.P. Suh, *Tribophysics,* Prentice-Hall, 1986

Basic Theory of Solid Friction

Jorn Larsen-Basse, National Science Foundation

Introduction

UNIVERSAL AGREEMENT as to what truly causes friction does not exist. It is clear, however, that friction is due to a number of mechanisms that probably act together but that may appear in different proportions under different circumstances. The recent introduction of sensitive and powerful techniques for measuring and modelling surfaces and even manipulating indicating surface atoms is creating a wealth of new information and is elucidating many previously unknown aspects of friction. Much still remains to be done, however, before a complete picture can emerge. In the meantime, this brief review of the various processes involved, as currently understood, is presented to familiarize the reader with the basic concept of friction and with the general approaches that can be used to control or minimize it.

The word "friction" is used to describe the gradual loss of kinetic energy in many situations where bodies or substances move relative to one another. For example, "internal friction" dampens vibrations of solids, "viscous friction" slows the internal motion of liquids, "skin friction" acts between a moving airplane and the surrounding air, and "solid friction" is

the friction between two solid bodies that move relative to one another. We are concerned here only with *solid friction,* which can be defined as "the resistance to movement of one solid body over another." The movement may be by sliding or by rolling; the terms used are "sliding friction" and "rolling fiction," respectively. Most of the discussion that follows deals with sliding friction.

The need to control friction is the driving force behind its study. In many cases low friction is desired (bearings, gears, materials processing operations), and sometimes high friction is the goal (brakes, clutches, screw threads, road surfaces). In all of these cases, constant, reproducible, and predictable friction values are necessary for the design of components and machines that will function efficiently and reliably.

It is useful to clearly separate the various terms and concepts associated with friction, such as "friction force," "friction coefficient," "frictional energy," and "frictional heating." These terms are defined below and in the "Glossary of Terms" in this Volume.

The *friction force* is the tangential force that must be overcome in order for one solid contacting body to slide over another. It acts in the plane of the surfaces and is usually proportional to the force normal to the surfaces, *N*, or:

$$F = \mu N \qquad \textbf{(Eq 1)}$$

The proportionality constants is generally designated μ or *f* and is termed the *friction coefficient.*

In most cases, a greater force is needed to set a resting body in motion than to sustain the motion; in other words, the *static coefficient of friction,* μ_s, is usually somewhat greater than the *dynamic* or *kinetic coefficient of friction,* μ_k.

A body on a flat surface will begin to move due to gravity if the surface is raised to the *friction angle,* θ, where:

$$\mu_s = \tan \theta \qquad \textbf{(Eq 2)}$$

See Fig. 1 in the article "Introduction to Friction" in this Volume.

To overcome friction, the tangential force must be applied over the entire sliding distance; the product of the two is *friction work.* The resulting energy is lost to heat in the in the form of *frictional heating* and to other general increases in the entropy of the system, as represented, for example, in the permanent deformation of the surface material. Thus, friction is clearly a process of energy dissipation.

Nature of Surfaces

Friction is caused by forces between the two contacting bodies, acting in their interface. These forces are determined by two factors besides the load; the *properties* of the contacting material and the *area* of contact. The friction forces are usually not directly predictable because both of these factors depend very much on the particular conditions. For example, the properties may be significantly different than expected from bulk values because the surface material is deformed, contains segregations, is covered by an oxide layer, and so on. Also, the real area of contact is usually much smaller than the apparent area of the bodies because real surfaces are not smooth on an atomic scale. Because of this close dependence of friction on the surface topography and on the properties of the surfaces and the near-surface layers, a brief discussion will be presented of the relevant characteristics.

Tabor (Ref 1) quotes W. Pauli: "God made solids, but surfaces were made by the Devil." Indeed, surfaces are extremely complicated because of their topography and chemical reactivity and because of their composition and microstructure, which may be very different from those of the bulk solid. Surface properties, composition, and microstructure may be very difficult to determine accurately, creating further complications.

Topography

The geometric shape of any surface is determined by the finishing process used to produce it. There will be undulations of wavelengths that range from atomic dimensions to the length of the component. These often result from the dynamics of the particular finishing process or machine used. There may be additional peaks and valleys caused by local microevents,

such as uneven deformation of hard microstructural constituents, local fracture, or corrosive pitting. Even after a surface has been carefully polished, it will still be rough on an atomic scale. It is useful to distinguish among macrodeviations, waviness, roughness, and microroughness (Ref 2) relative to an ideal flat surface (Fig. 1).

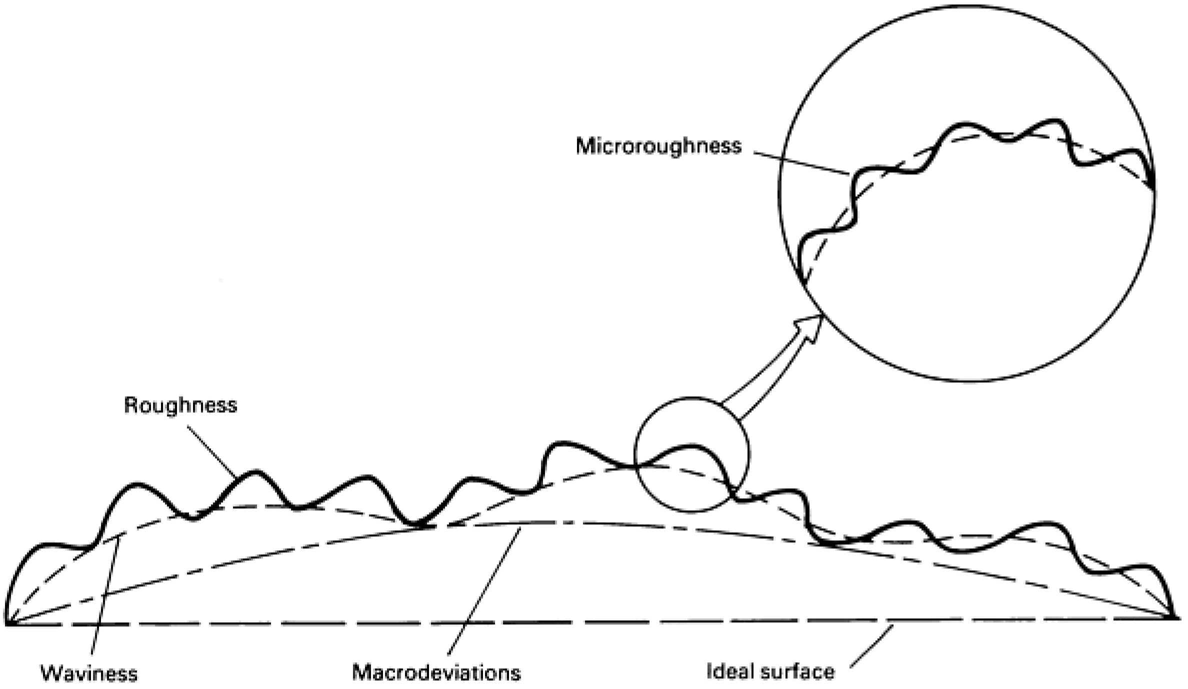


Fig. 1 Schematic showing selected types of surface deviations relative to an ideal solid surface

Macrodeviations are errors from--irregular surface departures from the design profile, often caused by lack of accuracy or stiffness of the machine system.

Waviness is periodic deviations from geometric surface, often sinusoidal in form and often determined by low-level oscillations of the machine-tool-workpiece system during machining (Ref 2). Typically, wavelengths range from 1 to 10 mm (0.04 to 0.4 in.) and wave heights from a few to several hundred micrometers (Ref 2).

Roughness is the deviations from the wavy surface itself, caused by geometry of the cutting tool and its wear, machining conditions, microstructure of the workpiece, vibrations in the system, and so on. Surface roughness changes as a surface goes through the wearing-in process, but may then stabilize.

Microroughness is finer roughness super-imposed on the surface roughness. It may extend down to the near-atomic scale and may be caused by internal imperfections in the material, nonuniform deformation of individual grains at the surface, or corrosion and oxidation processes that occur while the surface is being generated or during its exposure to the environment.

The peaks of surface roughness are called asperities. They are of primary concern in sliding friction and wear of materials, because these processes usually involve contacts between asperities on opposing surfaces or between asperities on one surface and asperity-free regions on the counterface. (The latter case may be unrealistic, but is often useful for modeling purposes.) Microroughness may affect the forces between surfaces, but has relatively little influence on surface deformation.

Roughness Measurement. A typical surface may have more than 10^5 peaks (Ref 3). Thus, it is generally not feasible to measure the height, shape, and location of every single peak on two matching surfaces in order to determine details of the contact. Instead, a simple profilometer trace is often used to measure and represent surface roughness. The stylus of the profilometer is a fine diamond with a fairly sharp tip, 2 μm or less in radius. It is drawn over the surface, and its vertical movement is amplified and recorded. The horizontal magnification is typically 100×, while the vertical magnification may vary from 500 to 100,000× (Ref 3), depending on the necessary resolution.

Because the stylus tip has a finite sharpness, it cannot shows very fine detail and tends to distort some shapes. For example, valley in the surface are shown narrower than they actually are and peaks are shown broader. Also, because only a fairly small portion of the surface can realistically be measured, the profilometer data are not absolute values and should be used only as relative data for comparison purposes. They are best used to compare surfaces produced by the same process--for example, by coarse and fine turning or by coarse and fine grinding.

Traditionally, the analog output of the profilometer is analyzed in terms of the deviation of the profile from the centerline. Two slightly different measures have been used. The roughness average, R_a, is the mean vertical deviation from the centerline and is the value most often used in Europe. The root mean square value, RMS, is the value most commonly used in the United States. It is calculated as the square root of the mean of the squares of the deviations and represents the standard deviation of the height distribution. Typical values for both roughness measures are 1.4 μm (55 μin.) for fine turned surface, 1.0 μm (39.4 μin.) for a ground surface, and 0.2 μm (7.9 μin.) for a polished surface (Ref 3). A table of typical values is given in Ref 2.

Other parameters used to measure roughness include skewness, R_{sk}; height, R_z; and bearing ratio curve.

Modern digitized instrumentation allows more detailed evaluation of the profilometer traces. It is now possible to scan a surface area by repeated but offset traces and to statistically evaluate the data for height distribution, asperity shape, and angle. Full use of the information available from modern instrumentation is still quite rare. The use of fractals to describe surface roughness has had limited success (Ref 4, 5), but much work remains to be done before it is clear whether this technique is more useful than traditional techniques. Additional information is available in the article "Wear Measurement" in this Volume.

Asperity Distribution Model. In contact situations, only the outer 10% of the asperities may be involved. Their height distribution can often be quite closely represented by the tail end of a Gaussian distribution (Ref 3). This distribution was used by Greenwood and Williamson (Ref 6) to derive an expression for elastic contact stresses. They also assumed that all of the asperities had the same tip radius. The Greenwood-Williamson (GW) model of surface roughness is commonly used to analyze contact mechanics of rough surfaces. It is probable, however, that the nature of the asperity height and shape distribution will change significantly once the surfaces begin to move against each other (Ref 7).

Composition

A surface is usually not completely clean, even in a high vacuum. Some of the events that can take place at surfaces are segregation, reconstruction, chemisorption, and compound formation (Fig. 2), as discussed in detail by Buckley (Ref 8).

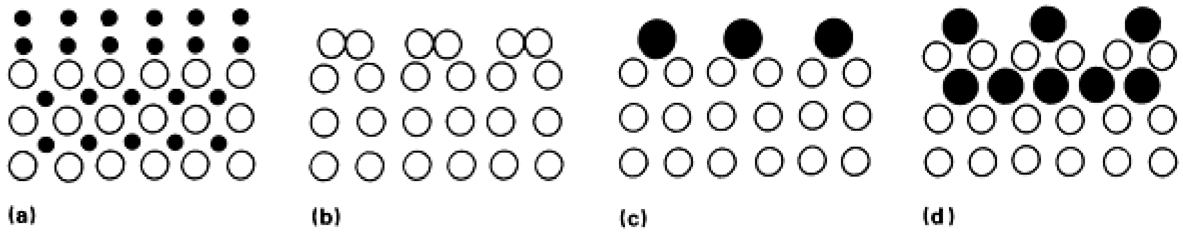


Fig. 2 Effect of composition on surface roughness defects. (a) Segregation. (b) Reconstruction. (c) Chemisorption. (d) Compound formation. Source: Ref 8

Segregation of alloy species to grain boundaries is a well-known phenomenon that may profoundly affect mechanical properties (Fig. 2a). Segregation to the surface may also take place. This generally occurs for small, mobile alloy or impurity atoms, such as interstitial carbon and nitrogen in iron, during processing or heat treatment. In some cases, the segregation of as little as 1 at.% of alloy element to the surface can completely dominate adhesion between contact surfaces (Ref 8). Significant changes in friction properties have been observed for ferrous surfaces with segregation of carbon, sulfur, aluminum, and boron, and for copper surfaces with segregation of aluminum, indium, and in (Ref 8). The nature of the changes friction due to surfaces segregation depends on the nature of the changes that the specific segregation in question causes in surface mechanical properties, adhesion, oxide film formation, and so on. For example, if certain metallic glasses containing boron are tested at increasing temperature, μ increases first with temperature, from

about 1.0-1.5 at room temperature to 1.8-2.5 at 350 °C (660 °F). Above 500 °C (930 °F), μ drops drastically (to about 0.25), a change that has been associated with the formation of boron nitride on the surface (Ref 8).

Reconstruction takes place when the outermost layers of atoms undergo a change in crystal structure (Fig. 2b). Examples include evaporation of silicon from a SiC surface upon heating, leaving behind a layer of carbon (Ref 8), and conversion of diamond surface layers to graphite or carbon during rubbing (Ref 9). Reconstruction may result in substantial changes in friction coefficient, but the fact that reconstruction has taken place may be evident only after careful characterization of the surface layers.

Chemisorption readily occurs on clean surfaces (Fig. 2c). Adsorbed species include water molecules from atmospheric moisture and carbon and carbon compounds also derived from the atmosphere or from lubricants used during operation or manufacture. The adsorbed species may also be components of various salts originating from the environment of from human handling of the component. The amount of adsorbed species, the degree of surface coverage, and the nature of the adsorbed molecule can substantially affect the adhesion between two surfaces, thereby directly or indirectly influencing friction behavior. For example, when a monolayer of ethane is introduced on a clean iron surface, the adhesive force drops from a value of greater than 400 dynes to 280 dynes (Ref 8). If the monolayer is acetylene, and force drops to 80 dynes. For a vinyl chloride monolayer, the force drops to 30 dynes--that is, to only 7 to 8% of the value for the clean surface.

Chemical compound formation may take place when surface comes into contact with a different solid, a gas, or a chemisorbed species. Without any tribological contacts, a surface will readily acquire a layer of oxide or hydroxide due to reactions with ambient moisture and oxygen. When two surfaces rub against each other, they may adhere at local spots that can reach elevated temperatures by frictional heating; interdiffusion may then take place, resulting in local compound formation in the surface layers (Fig. 2d). This can strongly affect friction. It is well known, for example, that friction between two metals that can form alloy solutions or alloy compounds with each other generally is greater than if the two are mutually insoluble. This fact has been used by Rabinowicz (Ref 10) to develop a generalized "map" showing which metals can safely slide against one another and which metal couples should be avoided (Fig. 3).

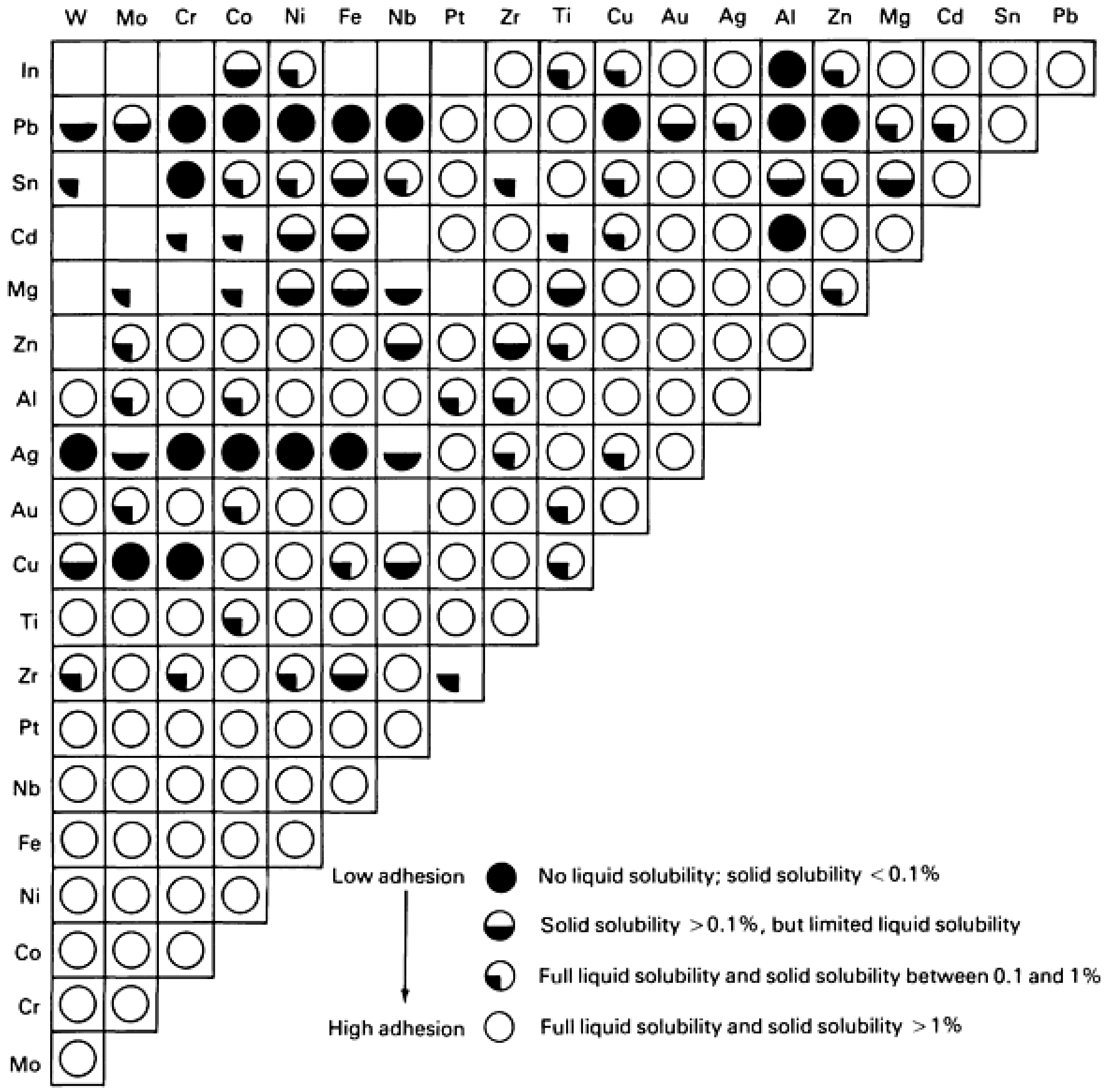


Fig. 3 Compatibility chart developed by Rabinowicz for selected metal combinations derived from binary equilibrium diagrams. Chart indicates the degree of expected adhesion (and thus friction) between the various metal combinations. Source: Ref 10

Surfaces rubbing against each other in the presence of organic compounds may catalyze the formation of polymeric layers, so-called tribopolymers, which may form more or less coherent layers on the surface. These can also affect friction behavior.

Mechanical compound formation is caused by the mechanical allowing of metallic wear particles and surface debris to form solid layers or segments of layers. A layer that forms preferentially on one of the sliding surfaces is often called a transfer layer (Ref 11). The wear particles involved in transfer layer formation are extremely small--of the size of dislocation cells in the heavily deformed surface layers of worm surfaces. These particles are pressed together with one another and with any other small particles present (oxides, oil-additive soaps, and so on) by the very localized, and therefore large, mechanical stresses that act on those asperities in contact with one another. The result is a more or less coherent, very thin transfer layer that may keep the surfaces from coming into direct contact with each other.

Transfer films also form when polymers or carbon rub against metal surfaces, but the formation mechanism may be somewhat different from that for metal-metal couples. The film forms gradually during the first 5 to 10 passes as polymeric or carbon wear particles adhere to the metal surface. The friction usually fluctuates during this stage; when the film is fully developed, the friction takes on a steady and usually low value.

Subsurface Microstructure

The layers immediately below the surface often have a microstructure that is different from the bulk. This is true for machined and ground surfaces, especially if the surface has been heavily worn. The surface layers of metals tend to become heavily deformed during wear, typically to a dept of deformation of about 40 μm (1575 μin.). Shear strains of 1100% and strain rates as high as 10^3/s have been estimated for the outermost layer (Ref 11). Because much of the deformation takes place in compression, otherwise brittle particles may be plastically deformed; for example, cementite lamellae in pearlite may be bent 90° with little or no cracking. The surface layers develop a very heavy dislocation concentration nd a subcell structure. The microstructural aspects of worn metallic surfaces have been reviewed in more detail by Rigney (Ref 11). Figure 4 illustrates some of the surface and subsurface features discussed above, primarily for metals.

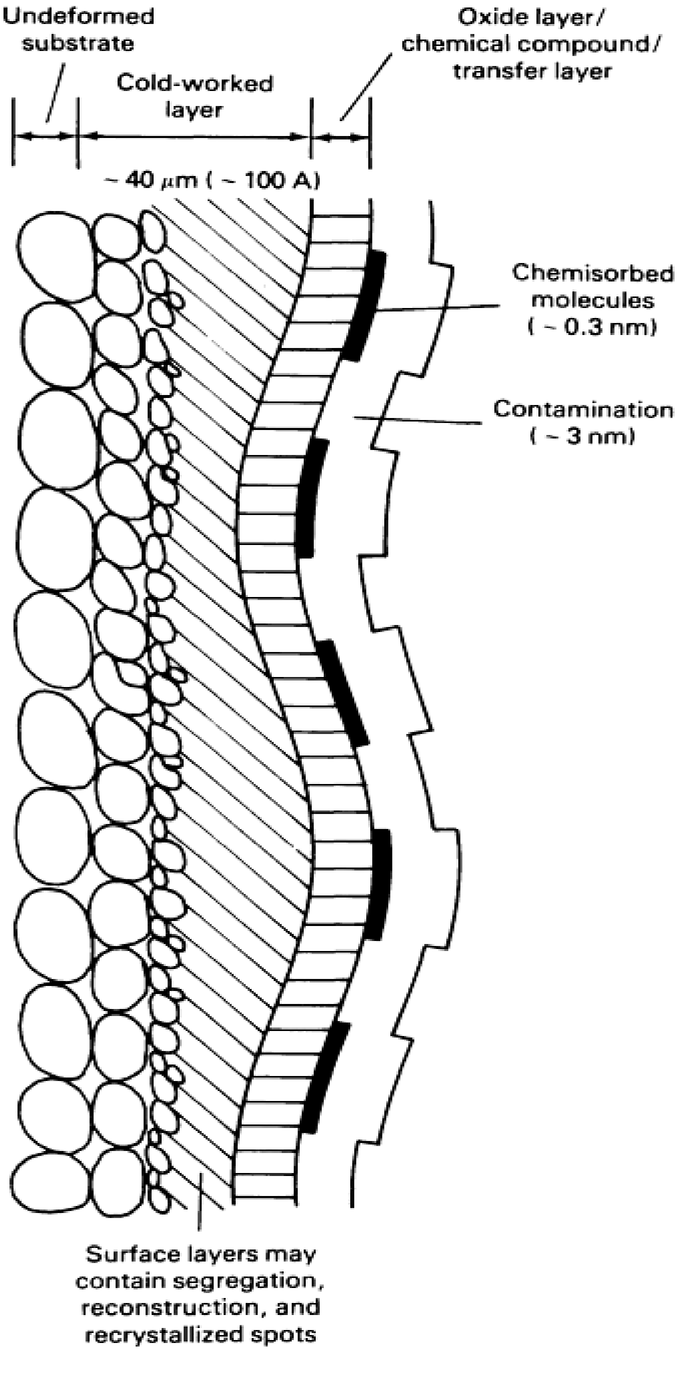


Fig. 4 Schematic showing typical surface and subsurface microstructures present in metals subject to friction and wear. Microstructures are not drawn to scale.

Friction under Lubricated Conditions

The nature, topography, and composition of the surface layers may be important also under lubricated conditions. Many sliding surfaces are lubricated to protect against war and to lower the friction. While most of the discussion here deals with dry sliding friction, it is instructive to briefly consider the transition between lubricated and dry conditions.

In a fully hydrodynamic situation, the lubricant film is sufficiently thick to keep the surfaces completely apart. The friction is then due to viscous dissipation within the lubricant and has little or nothing to do with the nature of the contacting materials. As the two surfaces are brought closer together, the asperities begin to come in contact and the zone of so-called "boundary lubrication" is entered. The degree of separation between the two surfaces can be measured by the ratio of the mean gap distance, *h*, to the composite roughness of the two opposing surfaces, σ. The composite roughness is defined by:

$$\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$$

where σ_1 and σ_2 represent the rms roughness of the two surfaces.

The h/σ ratio is often refereed to as the lambda (Λ) ratio. Generally, for surface whose height distributions are nearly Gaussian, if Λbecomes greater than 3 the conditions are full-film hydrodynamic conditions and asperity interactions are rare. For Λless than 3, asperity rubbing takes place and friction increases as h/σ decreases. If $\Lambda$is less than 1.5, surface deformation may take place and boundary lubrication conditions prevail (Ref 12). In this region, and as the gap is decreased further toward dry sliding, friction depends on what happens in a thin film of lubricant on the surfaces and at asperity contacts. Ideally, the surfaces would be separated by a lubricant film at all times. The ideal film would be one that has low shear strength between molecular layers parallel to the surface (and thus low friction), but which at the same time has strong bonds with the solid and thus prevents the opposing solids from coming into intimate contact with one another. The bonding is affected by the nature and composition of the surface layers; trace elements, such as sulfur in steel, can have significant effects on the formation of these films. Similarly, it is expected that new additive molecules will have to be developed as ceramic triboelements become more common, because the bonds formed with ceramic surfaces are quite different from those between currently used additives and metallic surfaces.

Basic Mechanisms of Friction

The specific physical, chemical, or materials-related microscopic events that cause friction are called the basic mechanisms of friction. A number of different mechanisms of this nature have been proposed over the past several hundred years, and each has had its proponents among scientists and engineers. Interestingly, the situation has changed relatively little, with some modifications, the same general basic mechanisms are still thought to be responsible for friction, and there is still a certain degree of partisanship regarding each mechanism. However, the general consensus seems to be that all the various mechanisms may be involved in the generation of friction but that dominant mechanism in each case depends on the particular situation. For the purpose of this discussion, friction is considered a systems property. It depends on the nature of the two surface, the materials, the environment, the application conditions, and certain characteristics of the apparatus, such as vibrations and specimen clamping.

The microscopic mechanisms that are involved, to varying degrees, in generating friction are (1) adhesion, (2) mechanical interactions of surface asperities, (3) plowing of one surface by asperities on the other, (4) deformation and/or fracture of surface layers such as oxides, and (5) interference and local plastic deformation caused by third bodies, primarily agglomerated wear particles, trapped between the moving surfaces (Fig. 5).

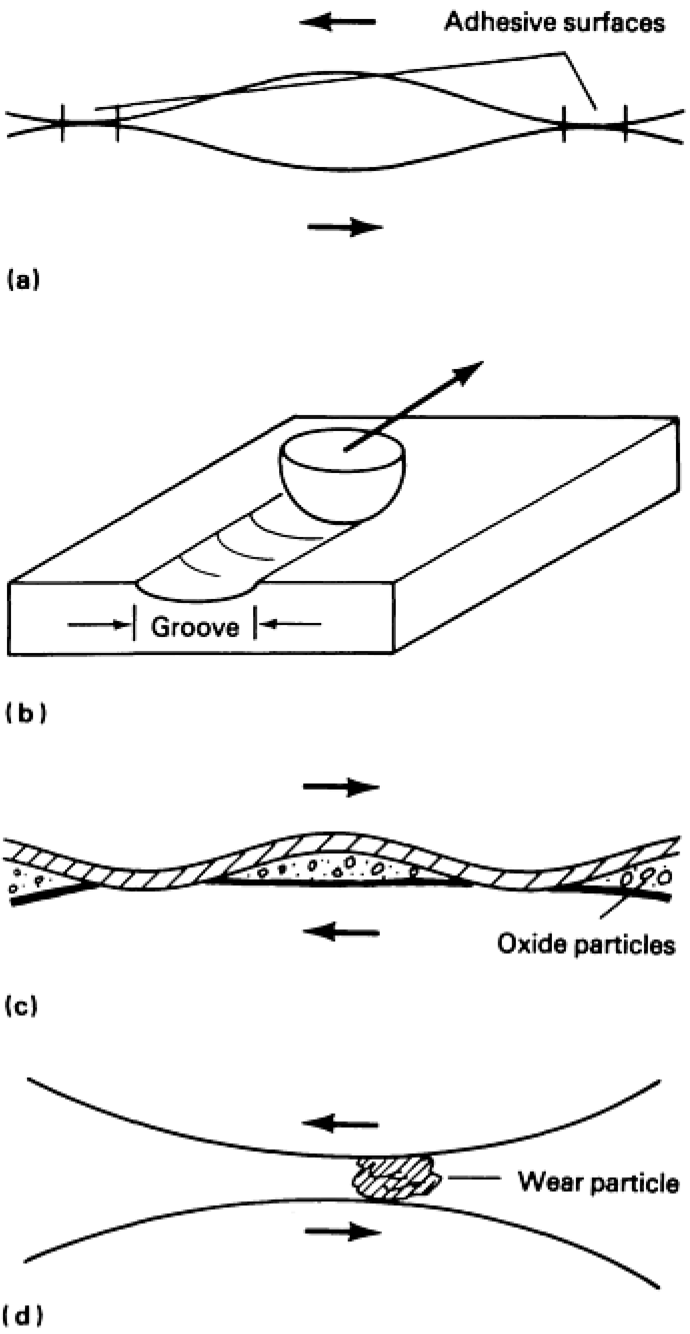


Fig. 5 Mechanisms on microscopic level that generate friction. (a) Adhesion. (b) Plowing. (c) Deformation and fracture of oxides. (d) Trapped wear particle

History

The history of the various attempts to scientifically explain friction has been described by Dowson (Ref 13) and by Bowden and Tabor (Ref 14) and has been briefly summarized by Ludema (Ref 15). The formative years of friction theory coincide with the general development of scientific thought during the 18th and 19th centuries. Basically, there were two schools of thought: a French school, which emphasized mechanical (elastic) interaction of surface roughness or asperities, and an English school, which emphasized "cohesion" or adhesion between the materials.

The French School began with a study published in 1699 by Amontons, who experimentally determined the two main "laws" of friction, often called Amontons' laws:

- The friction force is proportional to the applied load

- The friction force is independent of the apparent area of contact

The same relationships had been observed by Leonardo da Vinci 200 years earlier. Leonardo's studies were basically done before the world was ready for them, and this results were probably not known to the scientific world of Amontons' time. Leonardo's notes and manuscripts were hidden away in private collections and were discovered and printed fairly recently.

Amontons speculated that friction was caused by the interaction of surface roughness peaks. For hard surfaces, he envisioned that the asperities would be forced to slide up and down over one another; for more "elastic" materials, he suggested that the sliding would push aside the surface irregularity peaks.

The Swiss mathematician and theologian Euler, who gave us the symbols *e, i,* and π for common use in mathematics, elaborated on Amontons' theory from an analytical point of view. In 1750, while working in Berlin during a 25-year absence from his post as professor at St. Petersburg, he suggested that friction is caused by a ratcheting effect and that the friction work is the work to lift one body over the asperities of the other. The asperities would have a slope equal to or less than the friction angle. Euler developed the first clearly analytical approach to friction and treated it is an integral part of the mechanics of bodies in motion. He was also the first to use μ for the coefficient of friction and to draw a clear distinction between static and dynamic coefficient of friction, μ_k and μ_s (Ref 13).

The French physicist and engineer Coulomb confirmed Amontons' laws experimentally almost a hundred years after they were first expounded. In 1781, he suggested that friction was caused by mechanical interlocking of asperities and that the actual surface material on the individual asperities was functionless. Although his explanation was wrong, his name lives in quite prominently: the term "Coulomb friction" is still used for dry friction under most conditions (except where heavy plastic deformation is involved, as in metalforming).

The great contribution of the French school was to emphasize that contact occurs only at discrete points. Its major failing was its belief that the contact was determined solely by the original geometry of the asperities (Ref 14) and its exclusion of plastic deformation and asperity shape change from the model.

The English School was actually started by a Frenchman, Desaguliers, whose father was a Protestant priest who fled to England during a period of religious persecution. In a presentation to the Royal Society in 1724, Desaguliers introduced the concept of cohesive force (now called adhesion). He noticed that if two lead balls were pushed together with a light twist, they would stick together and that it took significant force to separate them again. He considered this cohesive force to be a universal phenomenon and suggested that friction can be largely attributed to the adhesion between asperities that come into intimate contact with one another.

Similar ideas were put forth by Tomlinson in 1929 and by Hardy in 1936; however, now they were based on the concept of molecular forces, which had been discovered in the interim and which are very short range in nature. Tomlinson even attempted to explain friction as a basic property derived from fundamental bonding forces working across the interface between the two metals in contact, combined with a partial irreversibility of the parallel force as atoms approach one another during sliding and then separate again.

Research in friction accelerated and reached a firm foundation with the work of Bowden and Tabor in the mid-20th century (Ref 16). Their work focused on adhesion as a major cause of friction, but also showed that more than the outermost layers is involved--that is, that both adhesion and deformation of the substance material are important contributors to the energy dissipation in friction. The adhesion theory of friction is often attributed to Bowden and Tabor, and, while they actually were not the first, they provided much supportive evidence; by including the plastic deformation of surface asperities, they showed that the mechanical properties of the surface material are also important.

In their early work, Bowden and Tabor assumed that the contacting asperities would deform to the point of plastic flow and reach a contact pressure equal to the indentation hardness of the material. The real area of contact, A_r, is then determined from:

$$A_r = \frac{N}{H} \quad \textbf{(Eq 3)}$$

where N is the normal load (in newtons) and H is the flow hardness (in N/m^2). If it is further assumed that friction is due to the shearing of bonds, then the friction force would simply be A_r times the relevant shear stress, τ. In that case:

$$\mu = \frac{F}{N} = \frac{A_r \tau}{A_r H} = \frac{\tau}{H} \quad \textbf{(Eq 4)}$$

This expression satisfies both of Amontons' laws in that contact area and load are eliminated. Because $H \sim 3\sigma_y$, where σ_y is the flow stress and $\tau \sim 0.5$ to $0.6\sigma_y$, a μ of 0.17 to 0.2 should result as a universal value for the coefficient of friction. Indeed, this value is often found for clean metals in air, but as later discovered, much higher values are found in a vacuum when the metals do not have a protective surface oxide film. It was suggested that shearing could also take place below one of the contacting asperities, especially if one of the materials was substantially weaker than the other. In that case, the weaker material would wear (Fig. 6).

Fig. 6 Schematic showing typical adhesive junction pull-off and wear generated by friction in the weaker of two materials

Tabor found qualitative support for the expression $F = A_r\tau$ by a simple experiment (Fig. 7). For the three pairs of slider versus flat:

- *Steel ball on indium flat:* μ= 0.6 to 1.2 because τ of the indium is low, but A_r is large
- *Steel ball on steel flat:* μ= 0.6 to 1.2, because τ is large, while A_r is small
- *Steel ball on steel flat with a thin indium coating:* μ= 0.06, because shearing both τ and A_r, are small. τ is small because shearing takes place in the indium, and A_r is small because the vertical load is supported by the steel substrate. The indium acts as a solid lubricant in this case

Tabor and his Cambridge students have continued work on friction and wear over the past half century. Much of our present understanding is due to their dedicated efforts.

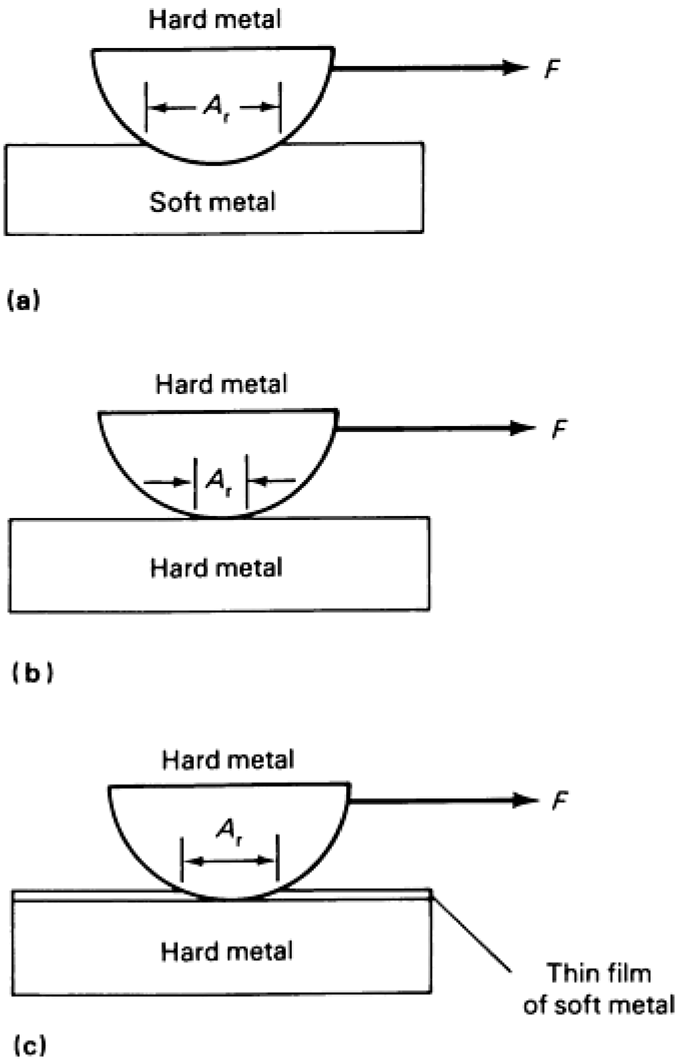


Fig. 7 Relation of friction force ($F = A_r\tau$) to metal substrate hardness. (a) Hard metal in contact with soft metal (small τ and large A_r). (b) Two hard metals of comparable hardness in contact with each other (large τ and small A_r). (c) Two hard metals of comparable hardness separated by a thin-film layer of soft metal deposited on one metal surface (both A_r and τ are small). Deposition of a thin film of a soft metal on a hard metal substrate yields the lowest friction force of the above-mentioned three cases. Source: Ref 16

The overview given in the following sections is not intended to be exhaustive, but rather to acquaint the reader with what many authorities in the field currently consider to be the mechanisms of friction. It is convenient to divide the discussion according to material type, with the understanding that there is considerable commonality among the groups and that most work to date has focused on metals.

Friction of Metals

Adhesion. The interfacial forces caused by adhesion dominate friction when the surfaces are very clean. The contacting surface asperities cold weld together and form intimate atomic bonds across the interface. This can take place at virtually not load, and because the size of the cold-welded area primarily depends on the smoothness of the surfaces and the closeness of their approach, A_r and F can be large. This means that μ can be 5, 20, 100, or even approach infinity. For the higher values, μ clearly loses its conventional meaning. Actually, recent work using molecular dynamics (Ref 17) and atomic force microscopy (Ref 18) has shown that when two surfaces are brought close together at a distance of a few atomic diameters, they will attract each other to form interatomic bonds. In this case, the normal force can be negative (a pull) which means that, strictly speaking, μ is negative. Again, in this situation the concept of friction has lost its conventional meaning.

From an engineering viewpoint, strong adhesion between sliding surfaces becomes important only for very clean surfaces in a very high vacuum. Adhesion problems have been studied extensively by Buckley (Ref 19), whose work has focused primarily on space applications, where the phenomena of adhesion and attendant seizure are extremely important.

High levels of friction in a vacuum environment were previously observed by Bowden and Tabor (Ref 16). They conducted an experiment to illustrate the effect on friction when the vacuum contains small amounts of molecular species that can chemisorb on the surface and thereby lower its tendency to form adhesive bonds. The results are shown in Fig. 8, which illustrates the behavior of pure iron with originally "clean" surfaces sliding against an identical specimen of pure iron. In the initial vacuum, friction was very high and seizure occurred. As oxygen entered the chamber at a low pressure of 10^{-4} mm Hg, μdropped to 2.3. If the oxygen pressure was increased 10-fold to 10^{-3} mm Hg, μdropped slightly to 2.1. It dropped further, to 1.9, as the oxygen pressure was increased to a few millimeters Hg. These μ-values are still very high compared with values that would occur under normal ambient conditions. However, leaving the surfaces exposed to a low oxygen pressure for a long period of time brought $\mu$to about 0.5, which is quite similar to values normally found.

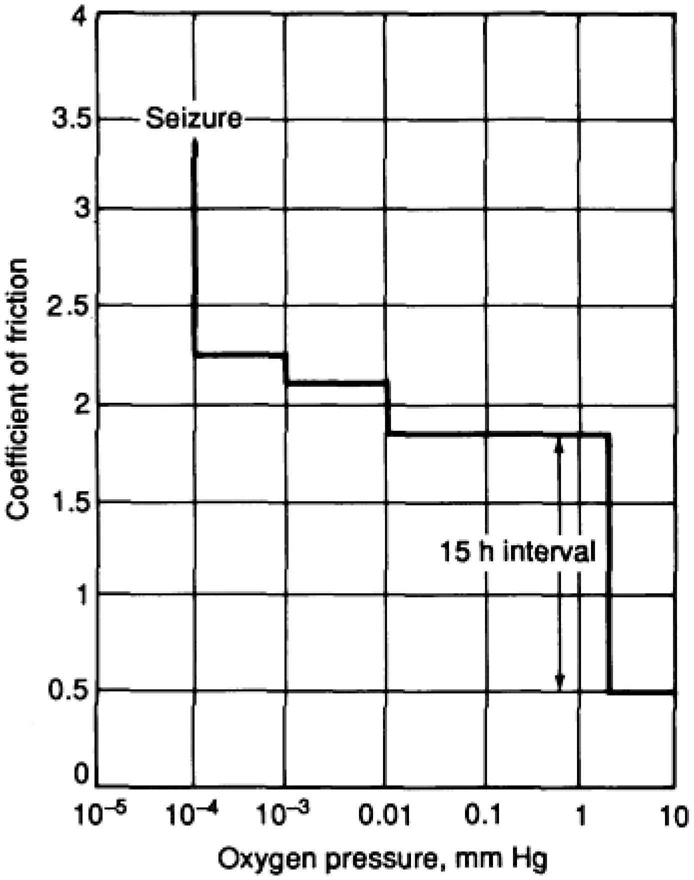


Fig. 8 Influence of oxygen on coefficient of friction of clean iron surfaces. Source: Ref 16

The behavior illustrated in Fig. 8 is not limited to iron sliding on iron, but is representative of the behavior of most metals when they are self-mated. Clean metal surfaces seize, or cold weld to each other, when they are brought together in a vacuum. When they are separated, chunks of material are usually transferred from one surface to the other, even for self-mated couples. When molecules that can adsorb to the surface, such as oxygen or water vapor, are admitted to the system, the friction drops because surface sites become covered with adsorbed atoms or even thin layers of oxide, and thus the surface area available for cold welding decreases. The more reactive a metal is, the more pronounced the effect.

Even materials that do not form oxides in the conventional sense exhibit this type of behavior. An example is diamond (Ref 19). In a vacuum, μ= 0.1, indicating that the diamond surface is protected by adsorbed species. If the two contacting surfaces are rubbed back and forth several hundred cycles in the same track, μincreases to 0.8 to 1, which indicates that the mechanical rubbing action has worn through the protective surface film (Ref 19).

In general, as mentioned previously, most classes of material have lower μ in air than in vacuum because of adsorbed molecules from the ambient air--water, oxygen, carbon dioxide, hydrogen, and so forth. One exception is soft glass, for which μ increases in ambient air. The reason for this behavior is that water molecules tend to chemisorb to the glass and bond the two surfaces together, thereby increasing their friction (Ref 19). If only dry air is used, the effect is not seen.

The amount of adhesion is also affected by the various other possible surface alterations discussed earlier: segregation of solute or impurity atoms to the surface, reconstruction of surface layers due to change in composition, and formation of compounds by chemical or mechanical action. The nature and the effects of all of these depend on the specific situation. For pure, film free surfaces in a vacuum, the adhesive friction depends on the size of the adhesion are, the strength of the adhesive bonds, and, in most cases, on the flow stress of the subsurface material, because that is usually where deformation takes place to accommodate the sliding.

Buckley (Ref 19) has demonstrated that the adhesion between two surfaces depends on the degree of matching between the crystal planes. The highest adhesion and friction forces are observed for matched planes of the same material. Lower values are found for matched planes of materials that are different but that have similar lattice dimensions and also show some mutual solubility. Still lower values are found when the two materials are insoluble in each other. For example, Buckley quotes high-vacuum values of $\mu = 21$ for the self-mated couple copper-copper; $\mu = 4$ for the closely fitting planes of the mutually soluble couple copper-nickel; $\mu = 2$ for copper-cobalt, where solubility exists but where one metal is face-centered cubic (fcc) and the other is hexagonal close-packed (hcp); and $\mu = 1.4$ for copper-tungsten, where no bulk solubility exists.

For matched planes and directions, the lowest values of μ in a vacuum are found for the planes with the highest atomic density--for example, the (111) planes in fcc metals or the basal plane in many hcp metals. These planes also have the lowest surface energy. Mismatched planes and directions yield lower values of μ.

Adhesive friction may also be related to other fundamental properties. One such property is the degree of ***d***-valence bond character of the transition metals (Ref 8, 19) (Fig. 9). Titanium, which has a very high degree of bond unsaturation, shows a strong tendency to bond with almost anything, such as a matching titanium surface or a nonmetal. As the degree of ***d***-bond character increases, the friction coefficient decreases--possibly because the greater the degree of bonding of a metal to itself, the less the bonding across the interface (Ref 19). There may be other possible explanations for this observed behavior. For example, it seems plausible that some or much of the effect could be due to changes in flow stress and flow behavior as the degree of ***d***-bond saturation changes.

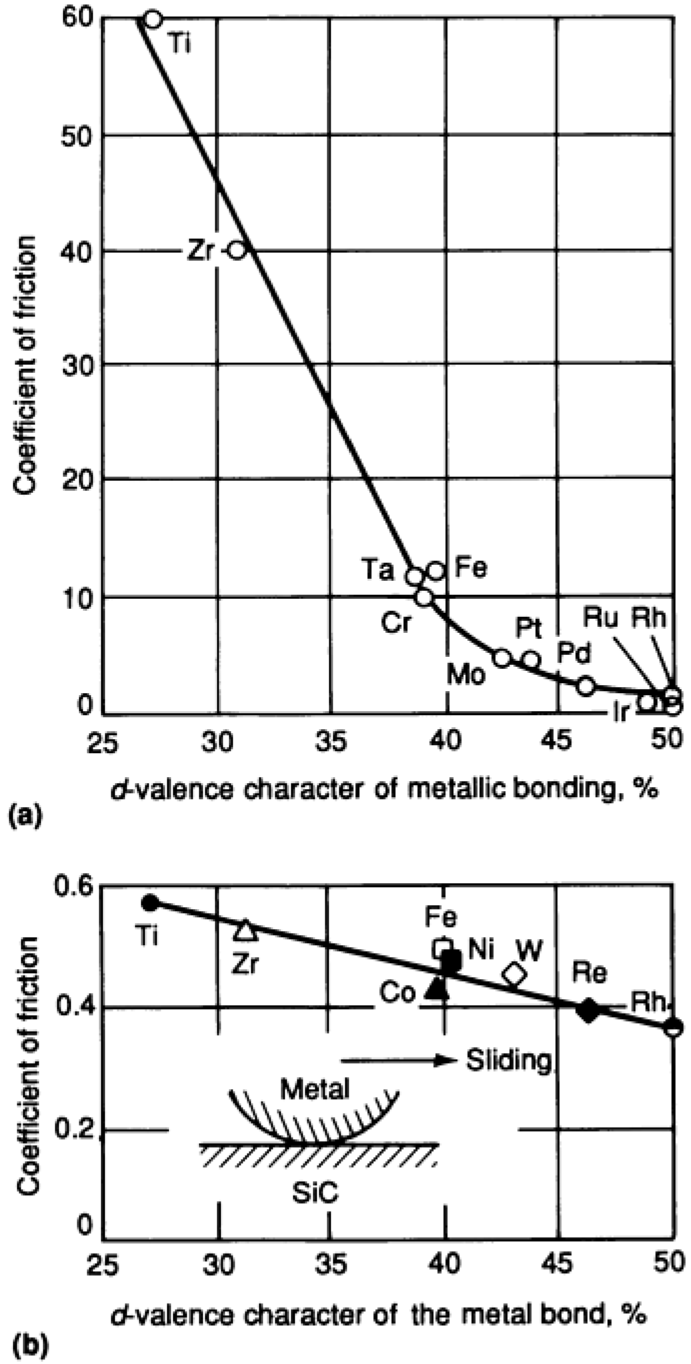


Fig. 9 Plot of coefficient of friction in a vacuum versus *d*-bond character of selected metals. (a) Metals in contact with themselves at very low load and sliding velocity. (b) Metals sliding in contact with single-crystal SiC. Source: Ref 19

Because plastic deformation is associated with friction, in most cases it is expected that even in a vacuum the flow stress of the material will affect μ. This is confirmed, for example, by results of friction tests with changing temperature for metals and alloys that undergo phase transformations. Figure 10 illustrates this for cobalt, which is hcp below 417 °C (783 °F) and fcc above this temperature. Below the transition temperature, μ= 0.35 for cobalt, because the basal hexagonal planes develop a preferred orientation, slips takes place between them, and there is little stain hardening. In this range, cobalt behaves like a solid lubricant. For the fcc structure above the transition temperature, μrises rapidly, possibly because of the significant work hardening of this structure.

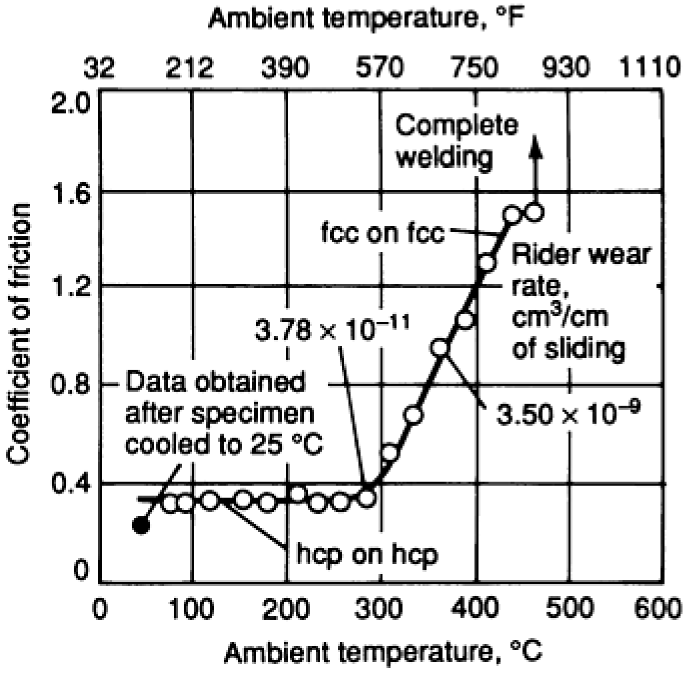


Fig. 10 Plot of coefficient of friction versus temperature as a function of phase transformation in a vacuum for cobalt sliding on cobalt. Sliding velocity, 1.98 m/s (6.5 ft/s). Source: Ref 19

Similar behavior has been observed for thallium, which also undergoes an hcp-fcc transformation. Not all hcp structures show low friction, however. A good example of this is titanium (Fig. 9). It appears that low friction occurs only for those structures that deform exclusively in the basal plane and that show low work hardening because the basal planes slide readily over one another. For lanthanum, which has three crystalline phases, the friction coefficients are in the order $\mu_{hcp} < \mu_{fcc} < \mu_{bcc}$ (bcc, body-centered cubic). Tin and tin-copper solid solutions show the opposite behavior with temperature: there is a drop in μ as the phase transformation from gray tin to white tin takes place upon heating 13 °C (55 °F). In this case, the behavior can also explained on the basis of the deformation properties of the two crystal structures. The high friction is exhibited by the low-temperature phase (gray tin). It has a diamond-type structure and exhibits a high degree of work hardening, while the white tin has a body-centered tetragonal (bct) structure that deforms more readily and that shows less work hardening and consequently lower friction (Ref 19).

In summary, adhesion is a very important component of friction in a vacuum. In extreme cases, it may lead to complete seizure of the two surfaces. The amount of adhesion depends on the nature of the surfaces, on their affinity for each other, and on their affinity for any adsorbates that may be present in the vacuum. The friction coefficient also depends on the flow stress properties of the near-surface material, because this is where deformation to accommodate the sliding most often takes place. In contrast, the effect of adhesion on friction under ambient conditions is more controversial, because it is not as unequivocally demonstrated.

Asperity Deformation. When asperities contact, they undergo elastic and plastic deformation to accommodate the stresses, as proposed by Bowden and Tabor (Ref 16), who divided the friction force into two components: a *plowing* term, F_p, and an *adhesion* term, F_a. The plowing term is due to energy dissipation in plastic deformation when the asperities interact with one another. If one metal is softer than the other, the hard asperities will produce visible grooves in the softer metal surface by a "plowing" action. For rubbers and polymers, the primary energy-dissipation mechanism is internal hysteresis. Most brittle materials can also undergo some plastic deformation in the compressive hydrostatic pressure region in front of and below a moving, plowing indenter or asperity. They also dissipate energy through microcracking.

The adhesion component of friction, F_a, is much more controversial, except when dealing with clean surfaces in a high vacuum, as discussed above. It is difficult to find a measurable force of adhesion between two "normal" engineering surfaces when they are pushed together under "normal" conditions. It has been suggested that this may be due to two factors:

- A large part of the surface is covered with films of oxides, adsorbates, and so forth, and only a few of the highest asperities are able to penetrate these films to form the metal-to-metal bonds needed for adhesion occur
- There is a large, elastically deformed region below the vary small, plastically stressed, adhering volume associated with the few spots of intimate contact. When the load moves, the elastic strain release overwhelms the adhesive bonds and ruptures them; consequently, significant adhesion is not seen in any force measurements

Bowden and Tabor originally suggested that that the reason a friction force would appear when two surfaces were slid parallel to each other, while almost no adhesion was seen if they were pulled apart without sliding, was that the junctions grew because of the horizontal forces and that friction was caused by the adhesion over these larger areas. This junction growth theory still has some supporters (Ref 18, 20), but it is now generally thought that adhesion does not contribute a clearly separate component to friction. Rather, adhesion is thought to be a component of the plastic deformation of asperities, a component that strongly influences the amount and the nature of the deformation. This approach has been reviewed by Johnson (Ref 21) and is discussed in some detail in his book (Ref 22).

The initial assumption that the contacts would be almost exclusively plastic was challenged by Archard (Ref 23), who pointed out that while it is reasonable to assume plastic flow for the first few traversals of one body over another, the same could not be assumed for machine parts that make millions of traversals during a life-time. The tallest asperities may flow plastically at first, but the surface must reach a steady state in which the load is supported elastically. For very rough surfaces, some initial plastic flow would certainly be expected, while for very smooth surfaces, the contact may be mostly elastic.

Greenwood and Williamson (Ref 6) attempted to model the condition at which changeover from elastic to plastic contact takes place. They used a multiasperity model and assumed that the asperities had a Gaussian height distribution and the same tip radius. They also assumed that the elastic deformation and stresses could be calculated from the Hertzian equations. Onset of plastic flow was taken to be the point where the maximum Hertzian pressure reaches about 0.6*H*, where *H* is the indentation hardness. This is in accordance with findings from studies of the indentation hardness of metals (see, for example, Ref 22).

It was convenient to introduce a *plasticity index:*

$$\psi = \frac{E'}{H}\sqrt{\frac{\sigma}{\beta}} \qquad \textbf{(Eq 5)}$$

where ψ is the plasticity index and E' is the effective elastic modulus from the Hertzian expressions, given by:

$$E' = \left(\frac{1-\nu_1^2}{E_1} + \frac{1-\nu_2^2}{E_2}\right)^{-1} \qquad \textbf{(Eq 6)}$$

where E_1 and E_2 are the elastic moduli and ν_1 and ν_2 are the Poisson ratios of the two materials. If one material is considerably harder than the other:

$$E' = \frac{E}{1-\nu^2} \qquad \textbf{(Eq 7)}$$

which can be called the "plane stress modulus." If the two materials are identical, E' is half of this.

In Eq 5, *H* is the indentation hardness (in N/m^2), σ is the standard deviation of the asperity height distribution, and β is the radius of the asperity tips. The plasticity index combines mechanical properties (E' and *H* and topographical properties (σ and β) of the solids in contact.

Greenwood and Williamson (Ref 6) found that while ψ may vary from 0.1 to 100 for real surfaces, in practice it falls in a narrow range. Typically, if $\psi > 1$, there is significant plastic flow; if $0.6 < \psi < 1$, there is some elastic and some plastic deformation; and if $\psi < 0.6$, plastic flow is unlikely.

Because load does not enter into the expression, it is clear that surface properties and surface topography, according to this model, play much greater roles in determining whether plastic deformation takes place. It is also clear that if the surface topography is such that plastic flow occurs initially, surface interaction in repeated passes may smooth the surface during run-in, until the standard deviation of the asperity height distribution (σ) decreases and/or the radius of curvature of the asperities (β) increases such that the plasticity index falls into the elastic range of contact.

Interestingly, in the GW model the real area of contract is almost proportional to load, even when the contact is entirely elastic. That is, Amontons' laws can be satisfied by elastic as well as by plastic contact conditions.

Plastic deformation may be of vital importance even when the total area of plastic contact is quite trivial. For example, on oxide-covered contacts, the plastic contacts will be the points where electrical and thermal conduction takes place, and they will also be the origin sites for much of the friction. However, many surfaces probably have no plastic contacts at all or have primarily elastic contacts rather than plastic contacts in each pass.

Somewhat more refined models of surface topography and of the deformation model have been proposed (Ref 24), as has a slightly different plasticity index (Ref 25). These are relatively minor improvements and have been omitted from this discussion for the sake of simplicity.

The effect of friction on the deformation process for idealized single-asperity contact has been described by Johnson (Ref 21) for the case of a two-dimensional asperity (a wedge), which lends itself to slip-line field analysis. The nature of the deformation under a blunt wedge depends on the interface friction between the wedge face and the surface of the metal (see Fig. 11 for a rigid, perfectly plastic material). If the friction is large, a cap of restrained material that does not flow plastically will form below the tip of the wedge. If the friction is zero, no such cap develops and the metal deformation takes place in a narrower zone around the indentation. If the wedge is made to move across the metal surface, it will initially dig deeper into the counterface, because the load will have to be supported on one side only. This is equivalent to the junction growth discussed above. Eventually, the wedge will return to the surface level, riding on its own "bow wave" and pushing a prow of plastically deformed material ahead of it. When the interface friction is high (perfect adhesion), the overall coefficient of friction approaches 1. If the adhesion is zero, pure plowing takes place. The wedge does not dig into the surface as deeply, and the final "bow wave" is much smaller (Fig. 12). In this case, the final friction is $\mu \sim \cot \alpha$, where α is the wedge half-angle.

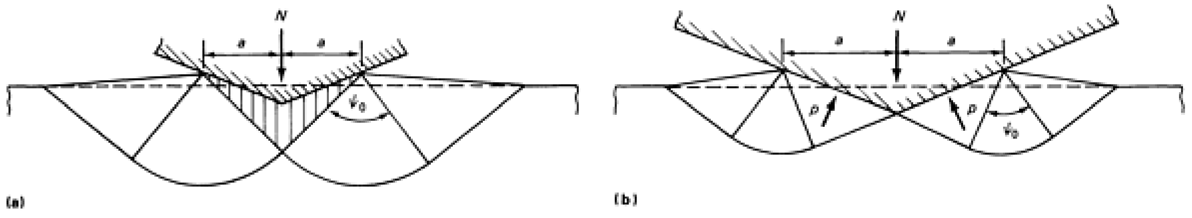


Fig. 11 Indentation of a rigid, perfectly plastic surface by a rigid blunt wedge. (a) With perfect adhesion, a cap of material adheres to the wedge face (shaded area). (b) With zero adhesion, the contact pressure, p_m, acts normal to the wedge face and the volume of deformed material is less. $p_m = 2k(1 + \psi_0)$. Source: Ref 21

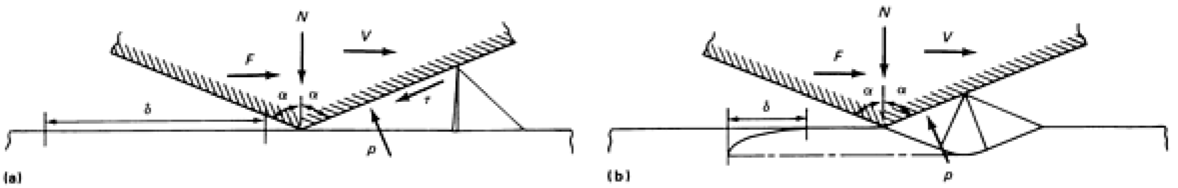


Fig. 12 Steady-state plowing of rigid, perfectly plastic material by a rigid wedge with half-angle, α, of 68°. (a) Perfect adhesion: τ= 0.73, *k* = 0.43*p*, μ= 1.0. (b) Zero adhesion : τ= 0, μ= cot α= 0.40. The plastic shearing of the surface relative to the bulk is shown by δ. Source: Ref 21

This idealized model illustrates the probable role of adhesion in the friction of ductile metals. Both with and without adhesion, the energy dissipation is caused by plastic deformation. Without adhesion, the plastic strains are relatively small , where μ= μ_p and is determined by the surface topography and generally has a value of less than 0.15. The presence of adhesion increase the plastic strains of plowing, causes substantial prow buildup, and yields $\mu\sim$1.0, irrespective of surface topography (Ref 21).

Challen and Oxley (Ref 25) have further investigated this model for various levels of interface strength, that is, for different levels of adhesion between the asperity face and the metal surface. Figure 13 shows their results for three different half-angles, 70°, 80°, and 90°, the last representing a flat punch. The friction coefficient, μ, is plotted versus the normalized interface strength, τ/*k*, where $\tau$is the interface shear strength and *k* is the shear yield stress of the metal. Relatively small changes in τcan change $\mu$from 1.0 to 0.5, a result well known from practical tests on metals in ambient air (see also Fig. 8). The μ-values at τ= 0 may be thought of as representing pure plowing and those at α= 90° as representing adhesion. When true adhesion takes place and there is no slip at the interface, μrises rapidly to a value of 1.0.

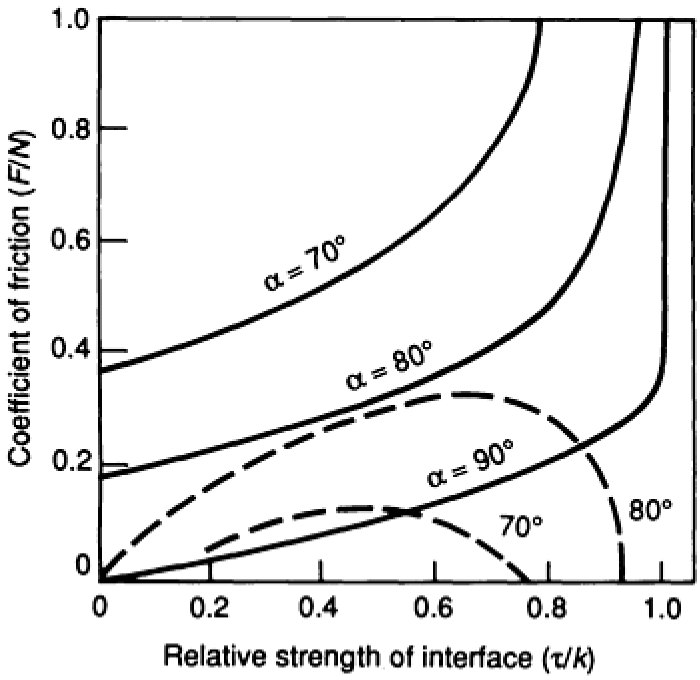


Fig. 13 Plowing by wedges with different half-angles and for various levels of partial adhesion. The "pure plowing" term, μ_p, is given by values of *F/N* when τ= 0; the "pure adhesion" term, μ_a, is given by *F/N* when α = 90°. The total friction coefficient, μ, is given approximately by μ_p + μ_a. The dashed line shows the fraction of total loss dissipated at the interface. Source: Ref 25

One must remember that conditions may be somewhat different in real life compared with the idealized conditions of plane strain indentation of elastic-perfectly plastic materials. The flow under a three-dimensional asperity is somewhat different from that under a wedge, and real surface material usually work hardens during the run-in period. However, the presentation based on the work of Johnson (Ref 21) and Challen and Oxley (Ref 25) illustrates the general phenomena, even if several minor details may be somewhat different under real-life conditions. One must also remember that the local contact areas may be at conditions quite different from those of bulk material. They may be heated by friction to temperatures that cause significant softening or even recrystallization and that may promote local oxidation. Furthermore, the local deformation happens rapidly and over a very short distance, while the resulting strain is large. Consequently, the associated strain rates can be very high. These conditions, coupled with the high hydrostatic pressure, may make it difficult to apply conventional constitutive equations for the material to the deformation that takes place in frictional contacts.

Deformation Energy. Rigney *et al.* (Ref 26, 27) have suggested that metallic friction can be determined on a deformation energy basis alone, from the strain energy that each pass of the slider introduces into the deformed layer. By equating the plastic work in surface deformation with the work done by the friction force, they derived the expression:

$$\mu = \frac{1}{1+n}\frac{\tau_0 \gamma_0}{H} \qquad \textbf{(Eq 8)}$$

where *n* is the work-hardening coefficient in the shear stress/shear strain flow equation:

$$\tau = \tau_0 \gamma^n \qquad \textbf{(Eq 9)}$$

and *H* is the hardness (in N/m^2).

This model has no adhesive component. Thus, current models of friction extend from pure adhesion to pure plastic deformation, with considerable coverage of the middle ground of adhesion plus elastic and plastic deformation. Although the modern models are much more detailed, their basic principles not very different from those of the early French and English schools. This is a tribute to the insight of the early pioneers in the field, but also underscores the complexity of the subject.

Third-Body Effects. An additional friction term based on mechanical deformation has been proposed by Suh (Ref 28, 29). He considers friction to have three components:

$$\mu = \mu_{asp} + \mu_{plow} + \mu_{part} \qquad \textbf{(Eq 10)}$$

where μ_{asp} is a contribution from deformation of the asperities, μ_{plow} is due to plowing effects, and μ_{part} is due to wear particles that remain in the wear zone and may agglomerate, work harden severely, and act as third bodies that deform the contacting surfaces.

The model was supported by findings such as that shown in Fig. 14. Here, the friction of a copper-copper sample rises from a rather low value as the distance of sliding increases (upper curve). If the surface is modulated to provide it with channels into which the particles can fall before they agglomerate and damage the surface, then the rise is not seen, and friction remains quite low (lower curve).

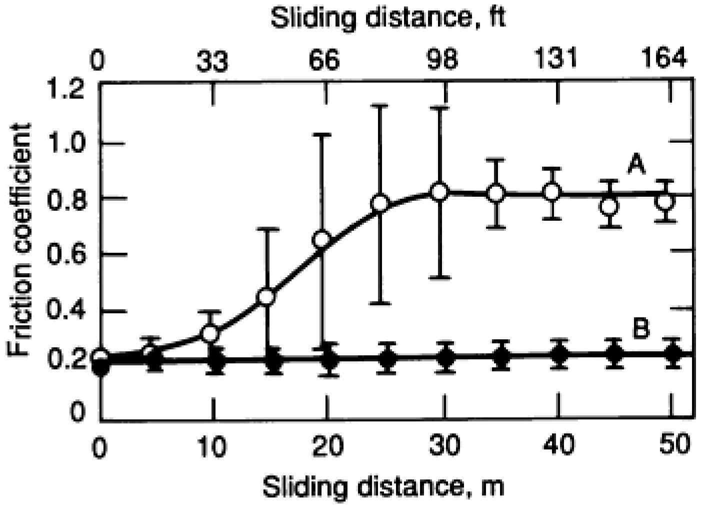


Fig. 14 Plot of coefficient of friction versus sliding distance for two types of copper surfaces. A, copper pin sliding on a copper flat. B, copper pin sliding on a modulated copper surface. Source: Ref 28

The role of trapped debris particles has also been recognized by Godet (Ref 30), who considers most wear behavior at steady state (for example, after 1000 cycles) to be controlled almost exclusively by third-body wear particles between the two surfaces. Adhesion and deformation in plowing and fatigue are still components of the entire process, but mainly because of their roles in formation of the wear particles. Interestingly, in this case friction is basically controlled by wear; the opposite statement--that wear is controlled by friction--is more often heard. Most likely, both statements are partially true.

Friction in metalforming is complex subject worthy of its own article an will be mentioned only briefly here. Because the contact loads during metalforming processes are high, the plastic zones beneath the asperities will merge and overlap and will eventually join with the deformation processes in the workpiece itself. Traditionally, it has been assumed that Coulombic friction controls the interface forces at low loads and that as the load grows to the point where the real area of contact is equal to the apparent area of contact, friction becomes independent of pressure and takes on the value *k,* which is the flow shear stress of the workpiece material (Fig. 15). It has been pointed out that *k,* is not a simple value; it is modified by the hydrostatic pressure and by geometric constraints, so that the final value under real conditions becomes somewhat lower than *k* determined from uniaxial tension tests (Ref 31). Because the situation is complicated by high stresses, high strain rates, frictional heating, surface oxides from heat treating, and so forth, the area of friction in metalworking is one of the least understood and most challenging for future work. An advanced analysis based on slip-line field studies has recently been presented by Kopalinsky *et al.* (Ref 32).

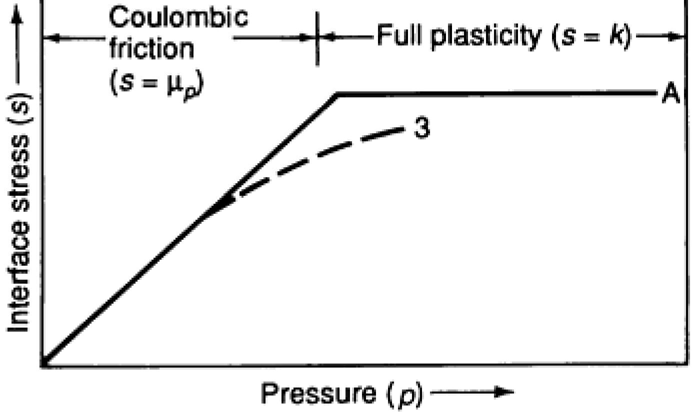


Fig. 15 Friction stress along the die surface in metalforming as a function of the normal pressure in the forming

process. A, ideal conditions. B, actual behavior

There is still some disagreement regarding the actual cause of friction of metals. In general, friction has several components:

$$\mu = \mu_a + \mu_p + \mu_e + \mu_{part} \qquad \textbf{(Eq 11)}$$

where

- μ_a is due to adhesion (or spot welding) between the surfaces. It is very important in high-vacuum applications and for very clean surfaces, and it may take over for certain metals that show seizure under ambient conditions. Under normal conditions, adhesion probably plays only a minor direct role, but it plays a significant indirect role by its effect on the plastic deformation in plowing
- μ_p is plastic deformation and plowing caused by deformation of one surface by hard asperities from the other. The result is formation of permanent grooves in the surface of the softer metal or pushing of a "bow wave" of material across the surface ahead of the indenter
- μ_e is a contribution from the elastic deformation of the material below the plastically deformed regions. It becomes more important as the surfaces are cold worked and smoothed during the run-in period
- μ_{part} is due to third-body particles trapped between the surfaces. These appear after some distance of sliding and are usually agglomerations of small wear particles. Their friction contribution is one of plastic deformation as they indent the surfaces or roll between them

Detailed information is available in the article "Appendix: Static and Kinetic Friction Coefficients for Selected Materials" in this Volume. It is clear that friction is compound property of the system in question and that prediction of friction from first principles is not yet possible.

Friction of Polymers

Friction in polymers is caused by many of the same mechanisms as for metals. There are other mechanisms, too, primarily because of differences in mechanical properties--in particular, the viscoelasticity, strain-rate sensitivity, and low thermal conductivity of polymers. In broad terms, friction is caused by mechanical deformation and surface adhesion, as for metals. The various friction mechanisms for polymers are illustrated in Fig. 16, 17, and 18 and are discussed in the following sections.

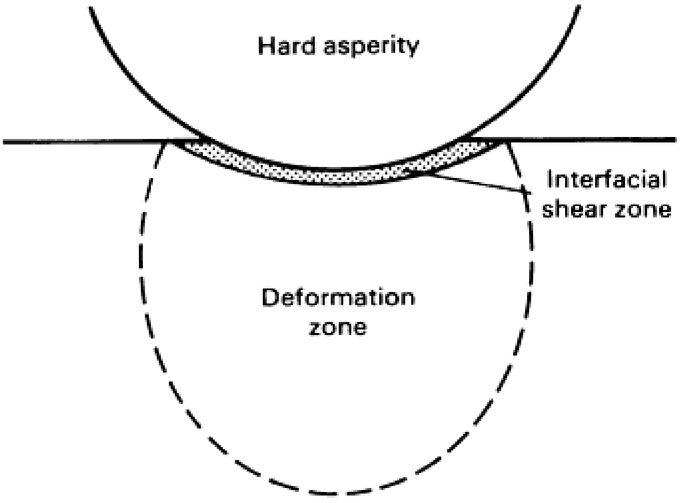


Fig. 16 Schematic showing a model of the friction dissipation zones present in a polymer. Source: Ref 33

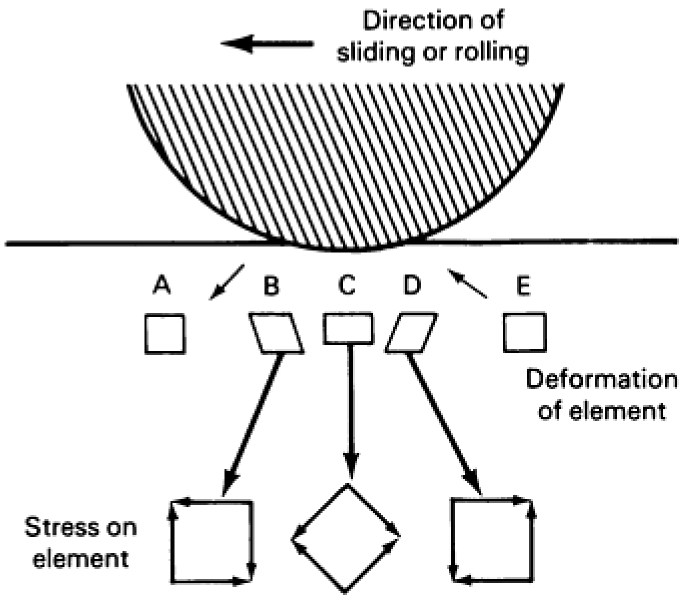


Fig. 17 A smooth, rigid sphere sliding or rolling over a viscoelastic material in the absence of surface adhesion. The progressive changes of stress beneath the contact are indicated for a chosen volume element. Each element undergoes about three cyclic deformations during the passage of the indenter. Source: Ref 33

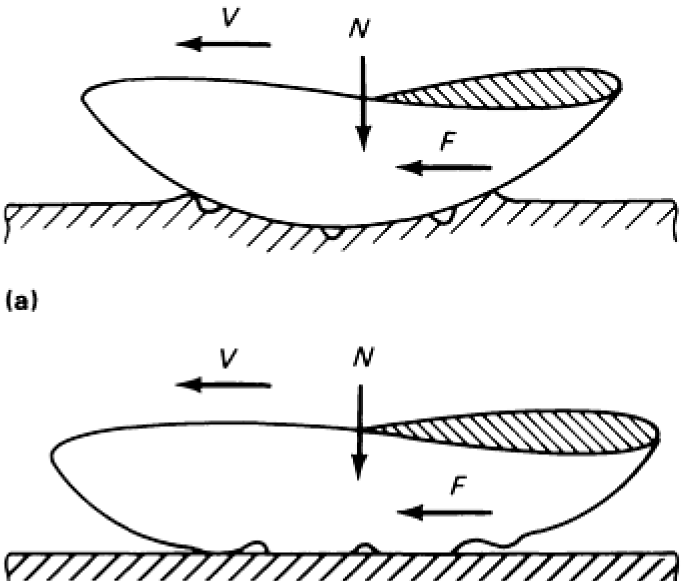


Fig. 18 Formation of Schallamach waves by buckling of the rubber surface at the compression side of contact. (a) For hard slider on rubber flat, waves move from front to back. (b) For rubber slider on hard flat, waves move from back to front. Source: Ref 21

Interfacial Zone Shear. Deformation takes place in a very thin layer (~100 nm) at the interface (Fig. 16). Slip may occur right at the interface itself, but is more commonly found within the polymer. In that case, a transfer layer forms on the other surface, possibly bonded adhesively via the carbon bonds (Ref 8) or held in place chemically by weaker interfacial bonds and mechanically by the surface roughness. Because the shear zone is very thin, it may experience extremely high shear rates with consequent local heating, exacerbated by the relatively low conductivity of the polymer matrix.

Some polymers do not appear to form transfer films, in which case sliding will take place at the interface between the two materials. This occurs primarily in highly cross-linked polymers and in some unlinked polymers below the glass transition temperature (Ref 33). Most semicrystalline polymers form transfer films, however. Typical examples at room temperature are polytetrafluoroethylene (PTFE) and low- and high-density polyethylene (HDPE). For PTFE, HDPE, and ultrahigh molecular weight polyethylene (UHMWPE), strain softening effects tend to lower the friction as sliding progresses. In general, for polymers whose frictional deformation takes place in the interfacial shear zone, friction is controlled by the flow stress at very high strains and strain rates and at somewhat elevated temperatures.

Deformation zone friction is friction caused by energy dissipation in deformation of the material below the indenter in the deformation zone (Fig. 16). This zone is similar in size to the plastically strained zone beneath an indenter on a metal substrate. Polymers have a characteristic ability to deform viscoelastically, in which case an imposed strain is fully recovered, usually with some time hysteresis. Organic polymers cover a full range of viscoelastic behavior, from essentially brittle to essentially ductile behavior, and the relative contribution of the two types of deformation behavior in any particular case depends strongly on deformation rate, temperature, and stress state (Ref 33). For glassy polymers, much of the energy dissipation in the deformation zone will be caused by microcracking, while ductile polymers produce grooving by permanent deformation. There is a transition between microcracking and plastic grooving that depends on sliding velocity, with lower speeds favoring more ductile deformation, as expected. Polymers in this general group include polymethyl methacrylate and polycarbonate.

Many polymers exhibit no permanent deformation after the indenter has passed, because they are viscoelastic and recover the original strain. The friction work is dissipated in the hysteresis of the deformation. Figure 17 shows the deformation patterns below a sliding (or rolling) hard contact. Each volume element undergoes about three cyclic deformations during the passage of the contact (Ref 33). The fraction of deformation energy lost (that is, the energy dissipation and thus the friction) is proportional to tan δ, where δ is the loss angle of the polymer at the deformation frequency. Corrections for local pressure and temperature effects may have to be made, because the conditions beneath the indenter are essentially adiabatic and quasi-hydrostatic pressure (Ref 33). If conditions are made more severe--for example, by increasing the interface friction or by increasing the depth of indentation--the deformation will change from viscoelastic grooving to tearing and plastic grooving. Additional information is available in the article "Solid Lubricants" in this Volume.

Friction of Elastomers

For fully viscoelastic materials, such as rubber and elastomers in general, interfacial friction appears to play an important role in dry sliding by strongly affecting the viscoelastic deformation behavior. Sliding involves generation of so-called Schallamach waves (Fig. 18). These waves are like giant dislocations that form on the compression side of the rubber member and then move through the contact zone toward the tension side. The rubber peels from the counterface and then becomes reattached. The reattachment energy is much smaller than the detachment energy (work of adhesion), so the latter dominates the behavior (Ref 21). The friction force, *F,* is:

$$F = (\Gamma)\, fA/V \qquad \textbf{(Eq 12)}$$

where Γ is the work of adhesion, *f* is the frequency of passage of Schallamach waves, *A* is the contact area, and *V* is the sliding velocity (Ref 21). Additional information is available in the article "Appendix: Static and Kinetic Friction Coefficients for Selected Materials" in this Volume.

Friction of Ceramics

The frictional behavior of ceramics is not well understood. In general, these materials might be expected to exhibit quite low friction, because the contributions from both adhesion and deformation are expected to be fairly low. The contribution from deformation should be relatively low because the hardness is high, and contribution from adhesion should be relatively low because formation of primary bonds across an interface between two ceramic materials requires the proper registry of positive ions on one side with negative ions on the other, and vice versa, a relatively rare condition. Nevertheless, ceramic materials exhibit quite high levels of friction--levels that change drastically as the wear mode changes. The friction is very strongly affected by the formation of surface films of oxides and hydroxides, which often have lubricating properties. Much work needs to be done in this area.

Other Materials

A number of materials exhibit special types of behavior that will be touched on briefly here. This behavior can be understood qualitatively from a detailed knowledge of the deformation behavior and surface chemistry of the materials in question. Additional information is available in the article "Appendix: Static and Kinetic Friction Coefficients for Selected Materials" in this Volume.

Graphite and MoS_2 exhibit low friction because of their layered structures. Each layer contains strong bonds that make it resistant to breakup and thereby enable it to carry substantial load. Weak bonds between the layers enable then to slide readily over one another. These materials and several materials with similar structures are used as solid lubricants.

Ice is a special material from a tribological point of view. Water is one of the very few materials that expands upon freezing. Consequently, ice can be made to melt by the application of pressure. Skates, skis, and sled runners generate their own lubricating film of water as they slide over the ice or snow. The lubricant forms at contacting asperities by a combination of frictional heating and pressure effects. If the ambient temperature drops sufficiently to strongly limit this melting (below about -25 to -30 °C, or -13 to -22 °F), operation of this equipment becomes difficult.

Rolling Friction

This article is primarily concerned with sliding friction, which is the friction that arises as one solid body slides over another. However, it has been known for thousands of years that it is easier to roll surfaces than to slide them. The resistance to rolling is called rolling friction and may be very low; for hard materials it may be as low as 0.001 (Ref 34). A very brief introductory discussion of this topic follows.

The use of rolling, as distinct from sliding, as a means of obtaining low coefficients of friction finds its greatest application in wheels and in ball and roller bearings. It is known that lubricants have little influence on rolling friction and that the resistance to "free rolling" (that is, rolling in the absence of an imposed tangential force) is made up of three components (Ref 22):

- Those arising from microslip and friction at the interface
- Those due to inelastic properties of the materials
- Those due to surface roughness

Microslip at the interface occurs both when the two bodies have different elastic properties and when their curvatures are different, but the effects are insignificant in both cases. Exceptions are found when there is a large area of contact, such as when a ball is rolling in a very deep groove, in which case the coefficient of rolling friction may approach 0.3 (Ref 34). Significant microslip may also occur when the rolling is tractive, that is, when large forces and moments are transmitted between the bodies through the contact zone. In this case the behavior will approach sliding.

Anelastic hyteresis losses dominate rolling friction in free rolling. If no permanent groove is formed from the rolling of a ball over a surface, the material beneath the front of the ball is compressed elastically and the material at the trailing part of the contact zone will expand elastic at the same time. If the material were ideally elastic, there would be no energy loss and the rolling friction would be zero (Ref 34). In reality, the deformation has some inelastic hysteresis and the corresponding energy loss is dissipated within the solids, at a depth corresponding to the maximum shear component of the shear stresses. If the thermal conductivity is low and the elastic hysteresis loss is high, this energy release can lead to failure by thermal stress beneath the surface (Ref 22).

The behavior of metals and ceramics in rolling contact is quite different from that of rubbers and polymers. For the former the anelastic hysteresis is governed by minute dislocation movements and is therefore usually very small, resulting in low rolling resistance. On the other hand, materials (such as rubbers) that exhibit full or partial viscoelastic deformation may have considerable rolling friction, and this friction may be quite sensitive to both temperature and deformation rate. Typical μ_r values for an automobile tire fall in the range of 0.01 to 0.03.

Surface roughness can affect rolling friction in two ways, although the affect usually is minor. First, for very lightly loaded rough hard surfaces, the energy expanded in lifting the body over the irregularities gives a small contribution to the rolling friction. Most of the energy transfer in this situation is by impact between surface irregularities, and the rolling friction due to this causes therefore increases with rolling speed (Ref 22). The second contribution arises from localized deformation. At the local asperities, the contact pressure may be concentrated to the point where permanent deformation

occurs, even if the bulk stress level is within the elastic limit. This can result in a decrease in rolling resistance with repeated traverses as the surface roughness is smoothed out by repeated plastic deformation of the protruding points (Ref 22).

In summary, rolling friction is usually very low, and it is primarily determined by anelastic deformation losses in the material. The greater the hysteresis loop of the deformation, the larger the energy loss in the deformation cycle and the greater the consequent coefficient of rolling friction.

Future Outlook

The basic mechanisms of friction are adhesion and mechanical deformation. Their relative roles are still the subject of much discussion. Frictional energy appears to be lost primarily as energy dissipates through deformation of the surface layers by elastic, plastic, and viscoelastic deformation and/or by microfracture of the surface material and possibly some mode II (shear) fracture of adhesive interface bonds. Adhesion is a primary cause of friction in high-vacuum environments and in instances of seizure. In most cases, however, surface films and contamination limit adhesion to a few small spots, where it can strongly influence the amount and nature of the local friction-generated deformation. Much work needs to be done to elucidate these basic mechanisms and to link minimum friction values (that is, those not determined by the system in question) with basic materials properties--if, indeed, that is possible.

References

1. D. Tabor, Status and Direction of Tribology as a Science in the 80's: Understanding and Prediction, *Tribology in the 80's,* Vol 1, NASA Conf. Publ. 2300, National Aeronautics and Space Administration, 1983, p 1-11
2. I.V. Kragelskii and N.M. Mikhin, *Handbook of Friction Units of Machines,* American Society of Mechanical Engineers, 1988 (transl. from Russian)
3. R.D. Arnell, P.B. Davies, J. Halling, and T.L. Whomes, *Tribology Principles and Applicants,* Springer-Verlag, 1991, p 1-66
4. F.F. Ling and S.J. Calabrese, On Fractal Dimensions of Engineering Surfaces, *Approaches to Modeling of Friction and Wear,* F.F. Ling and C.T.H. Pan, Ed., Springer-Verlag, 1987, p 167-173
5. A. Majumdar and B. Bhushan, Role of Fractal Geometry in Roughness Characterization and Contact Mechanics of Surfaces, *J. Tribology (Trans. ASME),* Vol 112, 1990, p 205-216
6. J.A. Greenwood and J.B.P. Williamson, Contact of Nominally Flat Surfaces, *Proc. R. Soc. (London) A,* Vol 295, 1966, p 300-319
7. J.A. Greenwood, Surface Roughness, *NATO Advanced Study Institute on Fundamentals of Friction,* 1991 (in press)
8. D.H. Buckley, Importance and Definition of Materials in Tribology: Status of Understanding, *Tribology in the 80's,* Vol 1, NASA Conf. Publ. 2300, National Aeronautics and Space Administration, 1983, p 19-44
9. C.S. Yust, C.J. McHargue, and L.A. Harris, Friction and Wear of Ion-Implanted TiB_2, *Mater. Sci. Eng.,* Vol A105/106, 1988, p 489-496
10. E. Rabinowicz, Determination of Compatibility of Metals Through Static Friction Tests, *ASLE Trans.,* Vol 14, 1971, p 198-205
11. D.A Rigney, Sliding Wear of Metals, *Ann. Rev. Mater. Sci.,* Vol 18, 1988, p 141-163
12. W.O. Winer and H.S. Cheng, Film Thickness, Contact Stress and Surface Temperatures, *Wear Control Handbook,* M.B. Peterson and W.O. Winer, Ed., American Society of Mechanical Engineers, 1980, p 81-141
13. D. Dowson, *History of Tribology,* Longman, 1979
14. F.P. Bowden and D. Tabor, *The Friction and Lubrication of Solids, Part II,* Clarendon Press, Oxford, 1964, Chapter XXIV
15. K.C. Ludema, Friction, A Study in the Prevention of Seizure, *ASTM Stand. News,* May 1987, p 54-58
16. F.P Bowden and D. Tabor, *The Friction and Lubrication of Solids, Part I,* Clarendon Press, Oxford, 1950

17. U. Landman, W.D. Luedtke, N. Burnham, and R.J. Colton, Atomistic Mechanisms and Dynamics of Adhesion, Nanoindentation and Fracture, *Science,* Vol 248, April 1990, p 454-461
18. Q. Guo, J.D.J. Ross, and H.M. Pollock, What Part Do Adhesion and Deformation Play in Fine-Scale Static and Sliding Contact?, *New Materials Approaches to Tribology: Theory and Application,* L.E. Pope, L.L. Fehrenbacher, and W.O. Winer, Ed., Materials Research Society, 1989, p 51-66
19. D.H. Buckley, *Surface Effects in Adhesion, Friction, Wear and Lubrication,* Elsevier, 1981, p 315-427
20. R.T. Spurr, The Role of Adhesion in Friction, *Friction and Traction,* Proceedings of 7th Leeds-Lyon Symposium on Tribology, D. Dowson, C.M. Taylor, M. Godet, and D. Berthe, Ed., Westbury House, London, 1981, p 34-38
21. K.L. Johnson, Aspects of Friction, *Friction and Traction,* Proceedings of 7th Leeds-Lyon Symposium on Tribology, D. Dowson, C.M. Taylor, M. Godet, and D. Berthe, Ed., Westbury House, London, 1981, p 3-12
22. K.L. Johnson, *Contact Mechanics,* Camprbridge University Press, 1985
23. J.F. Archard, *Proc. R. Soc. (London) A,* Vol 243, 1957, p 190
24. D.G. Evseev, B.M. Medvedev, and G.G. Grigoryan, Modification of the Elastic-Plastic Model for the Contact of Rough Surfaces, *Wear,* Vol 150, 1991, p 79-88
25. J.M. Challen and P.L.B. Oxley, An Explanation of the Different Regimes of Friction and Wear Using Asperity Deformation Models, *Wear,* Vol 53, 1975, p 229
26. D.A. Rigney and J.P. Hirth, Plastic Deformation and Sliding Friction of Metals, *Wear,* Vol 53, 1979, p 345-370
27. P. Heilman and D.A. Rigney, Sliding Friction of Metals, *Friction and Traction,* Proceedings of 7th Leeds-Lyon Symposium on Tribology, D. Dowson, C.M. Taylor, M. Godet, and D. Berthe, Ed., Westbury House, London, 1981, p 15-19
28. N.P. Suh, *Tribophysics,* Prentice-Hall, 1986
29. S.T. Oktay and N.P. Suh, Wear Debris Formation and Agglomeration, *J. Tribology (Trans. ASME),* Vol 114, 1992, p 379-393
30. M. Godet, Modeling of Friction and Wear Phenomena, *Approaches to Modeling of Friction and Wear,* F.F. Ling and C.T.H. Pan, Ed., Springer-Verlag, 1987, p 12-36
31. F.F. Ling and M.B. Peterson, Simple Model of Metalworking Friction Under Extreme Pressure, *Approaches to Modeling of Friction and Wear,* F.F. Ling and C.T.H. Pan, Ed., Springer-Verlag, 1987, p 83-88
32. E.M. Kopalinsky, X. Li, and P.L.B. Oxley, Modelling Tool-Work Interface Friction in Metal Working Processes, *Tribological Aspects in Manufacturing,* M.H. Attia and R. Komanduri, Ed., American Society of Mechanical Engineers, 1991, p 217-235
33. B.J. Briscoe, The Friction of Polymers: A Short Review, *Friction and Traction,* Proceedings of 7th Leeds-Lyon Symposium on Tribology, D. Dowson, C.M. Taylor, M. Godet, and D. Berthe, Ed., Westbury House, London, 1981, p 81-92
34. F.P. Bowden and D. Tabor, *Friction and Lubrication,* 2nd ed., Methuen 1967

Frictional Heating Calculations

Richard S. Cowan and Ward O. Winer, Georgia Institute of Technology

Introduction

WHEN COMPONENTS IN RELATIVE MOTION are mechanically engaged, the region of contact, which could be dry or separated by a lubricant film, experiences a temperature rise. This phenomenon has been the concern of numerous theoretical and experimental studies because its presence may affect the performance and longevity of the respective surfaces.

While such investigations yield differing effects dependent on the situation, the common observation of a temperature change has been associated with the presence of mechanical energy, which is required to overcome frictional resistance as sliding at the contact interface occurs. The energy, dissipated through conversion into thermal energy, is manifested as a temperature rise. At the microlevel, this increase can be substantial. A localized change in material properties, an enhancement in chemical reactivity, and ultimately, failure of the mechanical system can result.

Attempts to quantify temperature changes have led to the development of straightforward equations associated with the type of contact. Although obscured by such situational uncertainties as the coefficient of friction, real area of contact, time of heat source exposure, and the constancy of material properties, the computational methods outlined in this discussion are focused on the flash temperature; that is, the relative change between the surface temperature and bulk temperature of a component due to frictional energy dissipation at the surface. To a designer, such an analysis provides an indication of what temperature level to expect when surfaces are in contact, provided that the physical and chemical changes that may occur in a surface layer are accounted for.

Acknowledgement

The authors wish to express their appreciation to the George W. Woodruff School of Mechanical Engineering, at the Georgia Institute of Technology, for the sponsorship of this work.

Frictional Heating Nomenclature

In applying the developments associated with frictional heating, concepts have emerged that can dramatically affect the integrity of an analysis. The user needs to understand how such parameters as temperature, *T;* coefficient of friction, μ; heat partition factor, γ; heat source time, *t;* Péclet number, Pe; and real area of contact, A_r, are interpreted and how they contribute to the heat transfer model employed.

Bulk, Contact, and Flash Temperature. Simply defined as the average temperature of the body prior to frictional heating, the bulk temperature, T_b, remains constant in the body at some distance from the location of frictional energy dissipation. Upon frictional heating, the surface temperature ascends from this bulk temperature to a contact temperature, T_c, at each point comprising the real area of contact. This temperature increase is commonly referred to as the flash temperature, T_f. Therefore:

$$T_f = T_c - T_b \quad \textbf{(Eq 1)}$$

Some members of the engineering community regard flash temperature to be the absolute (actual) temperature of the contact spots. However, this discussion will not use this alternate usage of flash temperature.

Coefficient of Friction. The coefficient of friction, μ, is defined as the ratio of the tangential force required to move two surfaces relative to each other, to the normal force pressing these surfaces together. It is sensitive to a variety of factors, including:

- Material composition
- Surface finish
- Sliding velocity
- Temperature
- Contamination
- Lubrication
- Humidity
- Oxide films

As a result, for any two surfaces, μ may fluctuate over several orders of magnitude, varying with time and location. Because it enters the calculation of flash temperature to the first power, the coefficient of friction provides a major source of uncertainty.

Heat Partition Factor. When two surfaces engage in sliding over a given contact area, the thermal energy generated per unit time, Q, is assumed to be distributed such that part of the heat, namely $Q_1 = \gamma_1 Q$ penetrates body 1, as the remainder, $Q_2 = \gamma_2 Q$, enters body 2. The coefficients γ_1 and γ_2 are known as heat partition factors.

As a function of the thermal properties, bulk temperatures, and relative speeds of the respective components, expressions for γ_i have been developed recognizing that:

$$\gamma_1 + \gamma_2 = 1 \qquad \textbf{(Eq 2)}$$

and that the contact temperature at each point on the interface is identical for both surfaces. Typically, only the maximum or mean surface temperatures within a given contact area are equated for ease of analysis.

Heat Source Time. When a surface contact is exposed to frictional heating, an unsteady situation ensues because the temperature increase becomes a function of time as well as position. The size of the source as well as the thermal properties and speed of the respective materials determine the transient behavior.

The surface receives thermal energy only for the time, t, that the heat source exists. Gecim and Winer (Ref 1) estimate that for a circular contact of radius a, a steady-state temperature is reached in a time such that the Fourier modulus, F_0, of each surface reaches 100. The Fourier modulus is a dimensionless heat transfer grouping:

$$F_0 = \frac{D_i \cdot t}{a^2} \qquad \textbf{(Eq 3)}$$

where D_i is the thermal diffusivity of body i.

Péclet Number. The Péclet number, Pe, is a dimensionless heat transfer grouping defined by

$$\text{Pe} = \rho c_p V L_c / k \qquad \textbf{(Eq 4)}$$

where ρ is the density; c_p is the specific heat at constant pressure; V is the velocity; L_c is a characteristic length; and k is the thermal conductivity (Ref 2). It relates the thermal energy transported by the movement or convection of the medium, to the thermal energy conducted away from the region where the frictional energy is being dissipated.

In computing the Péclet number, L_c is usually expressed as a measure of the contact dimension in the direction of sliding motion. Thus, for relationships presented in this review, the Péclet number can be expressed as:

$$\text{Pe}_i = \frac{V_i \cdot L_c}{D_i} \qquad \textbf{(Eq 5)}$$

where V_i is velocity of surface i, tangential to contact (in m/s); L_c is the contact width for line contact (w), or contact radius for circular contact (a), in the direction of motion (in m); and D_i is the thermal diffusivity of material i (in m^2/s), which is given by:

$$D_i = \frac{k}{\rho \cdot c_p} \qquad \textbf{(Eq 6)}$$

Real Area of Contact. At the microlevel, it is observed that a seemingly smooth surface is composed of a series of asperities. Thus, when contact between surfaces is made under low pressure, the interface is not one coherent area, as assumed by Hertz (Ref 3), but is made up of several small regions where respective surface peaks touch. Generally, the real area of contact (A_r) is only a fraction of the apparent area or contact patch. The real area is considered to be proportional to the normal load, and inversely proportional to the hardness of the softer of the two contacting materials.

The size of and distance between real contact areas within the apparent area will affect the distribution of thermal energy as variations in localized pressure and interactions from neighboring contacts result. This may have little effect on the average apparent contact surface temperature (given that the actual bearing area is small); however, the maximum temperature change from frictional heating at an asperity contact can be significantly, as evidenced by visible hot spots (Ref 4).

Idealized Models of Sliding Contact

Frictional heating calculations are developed on the premise that thermal energy is generated at an area of real contact, and that the energy is conducted away into the bulk of the rubbing members. Thus, the theory requires solution to equations for the flow of heat into each body in such a proportion as to yield equivalent surface temperatures over the contact region.

The following analyses demonstrate the formation of idealized models to represent sliding contact. Expressed in terms of the size of a heat source, the rate of heat flow, and the velocity and properties of the materials in contact, the computations obtained can be used in approximating practical conditions of a similar nature.

General Contact Analysis. The sliding contact may be considered as two solid bodies, of which one or both move at uniform speed past a band-shaped heat source (Fig. 1). This source has a heat flux distribution, *q*, with an average value of q_{av}.

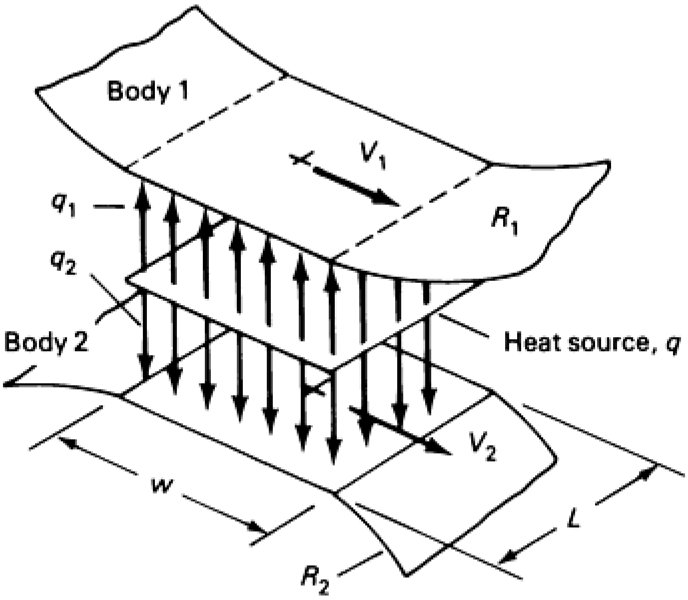


Fig. 1 Schematic showing key parameters that affect heat distribution in an ideal sliding contact model. V_1 and V_2, velocities of surface 1 and surface 2, respectively, both velocities being tangential to contact and normal to contact length; *q*, heat flux distribution; q_1 and q_2, portion of heat distribution that penetrates surface 1 and surface 2, respectively; R_1 and R_2, radius of curvature of surface 1 and surface 2, respectively; *L* and *w*, length and width, respectively, of heat source

The maximum contact temperature, T_c, will occur at the surface of either body. It may be simply computed from:

$$T_c = T_{f_i} + T_{b_i} \qquad \textbf{(Eq 7)}$$

where T_{f_i}, the maximum flash temperature of body *i* (in °C), is superimposed on its respective bulk temperature, T_{b_i}.

Considering that one or both bodies will move relative to the heat source, the maximum flash temperature (in °C) for a moving surface has been related to the parameters described as:

$$T_f = F \cdot \frac{\gamma \cdot q_{av}}{b} \cdot \sqrt{t}$$ **(Eq 8)**

where t represents the time during which any point on the surface is exposed to heat (Ref 5). Variable b denotes a thermal contact coefficient equal to the square root of the product of the specific heat (c), density (ρ), and thermal conductivity (k) of the material. Coefficient F is dependent on the form of the heat flux distribution, q, over the width of the heat source. For a square heat source with uniform distribution in which $q = q_{av}$, F is equivalent to $2/(\pi^{1/2}) = 1.13$, closely approximating a semielliptical distribution where $F = 1.11$ (Ref 6). Furthermore, the product $\gamma \cdot q_{av}$ represents the portion of heat entering the body where γ is the heat partition factor. The average total heat flux, q_{av} (in W/m^2), generated by friction between the two loaded surfaces can be expressed as:

$$q_{av} = \mu \cdot p_{av} \cdot V_r$$ **(Eq 9)**

where μ is the coefficient of friction, p_{av} is the average pressure according to Hertzian contact theory, and V_r denotes the relative sliding velocity between the two surfaces.

Should one of the bodies of Fig. 1 be stationary or moving such that there is sufficient time for the temperature distribution of a stationary contact to be established, the maximum flash temperature of the body is determined as a function of its thermal conductivity, k, and the heat flux distribution, q. Table 1 summarizes an assortment of maximum flash temperature relationships for point and line contacts based on a uniform heat flux distribution, $q = q_{av}$.

Table 1 Selected maximum flash temperature relationships for line and point contacts based on uniform heat flux distribution

Type of contact	Surface relative to heat source	Maximum flash temperature, T_f(a)
Line (L, length; w, width)	Stationary	$\frac{Q}{\sqrt{\pi}Lk}$
	Moving	$1.13 \frac{Q}{Lk} \sqrt{\frac{D}{Vw}}$
Circular (a, radius)	Stationary	$\frac{Q}{\pi ak}$
	Moving	$0.508 \frac{Q}{ak} \sqrt{\frac{D}{Va}}$

(a) Q, thermal energy generated per unit time; k, thermal conductivity; D, thermal diffusivity; V, velocity

Examples of frictional heating, with both bodies or only one body in motion, are discussed in the sections "Line Contact Analysis with Two Bodies in Motion" and "Circular Contact Analysis with One Body in Motion" in this article.

Line Contact Analysis with Two Bodies in Motion. Using the model of Fig. 1, the total heat flux developed from the two moving bodies in line contact must be partitioned (see the section "Heat Partition Factor" in this article). Blok estimated this by equating the maximum flash temperature of each surface using Eq 8, assuming equivalent bulk temperatures and a time of contact, t, equal to w/V_i, where w is the width of the heat source and V_i is the velocity of body i (Ref 6). As a result, the portion of heat withdrawn by surface 1 is:

$$\gamma_1 = \frac{n}{n+1}$$ **(Eq 10a)**

where:

$$n = \frac{b_1 \sqrt{V_1}}{b_2 \sqrt{V_2}} \qquad \textbf{(Eq 10b)}$$

As the two components have equal bulk temperatures T_b, the contact surface temperature (in °C) of either body can be simply expressed as:

$$T_c = T_f + T_b \qquad \textbf{(Eq 11)}$$

Should both bodies be moving such that the Péclet number of each surface is at least 2, the combination of Eq 8 and 9 for use in Eq 11 yields T_f (in °C):

$$T_f = 1.13 \cdot \mu \cdot \frac{W}{L\sqrt{w}} \cdot \left(\frac{|V_1 - V_2|}{b_1 \sqrt{V_1} + b_2 \sqrt{V_2}} \right) \qquad \textbf{(Eq 12)}$$

where μ is the coefficient of friction, w is the contact width (in m), L is the contact length (in m), W is the normal contact load (in N), V_1 and V_2 are velocities (in m/s) of surfaces 1 and 2, tangential to contact and normal to contact length, respectively, and b_1 and b_2 are thermal contact coefficients of bodies 1 and 2 ($W \cdot s^{1/2}/m^2 \cdot °C$); in which b_i is:

$$b_i = \sqrt{k_i \bullet \rho_i \bullet c_i} = k_i/\sqrt{D_i} \qquad \textbf{(Eq 13)}$$

Note that k_i, ρ_i, c_i, and D_i are the thermal conductivity, density, specific heat per mass, and thermal diffusivity of body i, respectively. Representative values of common materials are found in Table 2.

Table 2 Typical thermal properties of common materials

Material	Properties at 20 °C (70 °F)				Thermal conductivity (*k*) at selected temperatures, W/m · °C									
					Temperature, °C (°F)									
	Density (ρ), kg/m^3	Specific heat (*c*), kJ/kg · °C	Thermal conductivity (*k*), W/m · °C	Thermal diffusivity (*D*), (m^2/s) × 10^5	-100 (-148)	0 (32)	100 (212)	200 (392)	300 (572)	400 (752)	600 (1112)	800 (1472)	1000 (1832)	1200 (2192)
Aluminum														
Pure	2707	0.896	204	8.418	215	202	206	215	228	249	. . .	. . .	. . .	. . .
Al-Cu (Duralumin): 94-96% Al; 3-5% Cu; trace Mg	2787	0.883	164	6.676	126	159	182	194	. . .	. . .	. . .	. . .	. . .	. . .
Al-Si (Silumin): 86.5% Al; 1% Cu	2659	0.867	137	5.933	119	137	144	152	161	. . .	. . .	. . .	. . .	. . .
Al-Si (Alusil): 80% Al; 20-22% Si	2627	0.854	161	7.172	144	157	168	175	178	. . .	. . .	. . .	. . .	. . .
Al-Mg-Si: 97% Al; 1% Mg; 1% Si	2707	0.892	177	7.311	. . .	175	189	204	. . .	. . .	. . .	. . .	. . .	. . .
Lead	11,373	0.130	35	2.343	36.9	35.1	33.4	31.5	29.8	. . .	. . .	. . .	. . .	. . .
Iron														
Pure	7897	0.452	73	2.034	87	73	67	62	55	48	40	36	35	36
Wrought, 0.5% C	7849	0.46	59	1.626	. . .	59	57	52	48	45	36	33	33	33
Steel:														
0.5% C	7833	0.465	54	1.474	. . .	55	52	48	45	42	35	31	29	31
1.0% C	7801	0.473	43	1.172	. . .	43	43	42	40	36	33	29	28	29
1.5% C	7753	0.486	36	0.970	. . .	36	36	36	35	33	31	28	28	29
Nickel steel:														
0% Ni	7897	0.452	73	2.026	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
20% Ni	7933	0.46	19	0.526	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
40% Ni	8169	0.46	10	0.279	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
80% Ni	8618	0.46	35	0.872	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
Invar: 36% Ni	8137	0.46	10.7	0.286	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
Chrome steel:														
0% Cr	7897	0.452	73	2.026	87	73	67	62	55	48	40	36	35	36
1% Cr	7865	0.46	61	1.665	. . .	62	55	52	47	42	36	33	33	. . .
5% Cr	7833	0.46	40	1.110	. . .	40	38	36	33	29	29	29	29	. . .
20% Cr	7689	0.46	22	0.635	. . .	22	22	22	22	24	24	26	29	. . .
15% Cr, 10% Ni	7865	0.46	19	0.526	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
18% Cr, 8% Ni	7817	0.46	16.3	0.444	. . .	16.3	17	17	19	19	22	26	31	. . .
20% Cr, 15% Ni	7833	0.46	15.1	0.415	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
25% Cr, 20% Ni	7865	0.46	12.8	0.361	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
Tungsten steel:														
0% W	7897	0.452	73	2.026	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
1% W	7913	0.448	66	1.858	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
5% W	8073	0.435	54	1.525	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .

10% W	8314	0.419	48	1.391	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
Copper														
Pure	8954	0.3831	386	11.234	407	386	379	374	369	363	353	. . .	. . .	. . .
Aluminum bronze: 95% Cu; 5% Al	8666	0.410	83	2.330	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
Bronze: 75% Cu; 25% Sn	8666	0.343	26	0.859	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
Red brass: 85% Cu; 9% Sn; 6% Zn	8714	0.385	61	1.804	. . .	59	71	. . .	. . .	. . .	. . .	. . .	. . .	. . .
Brass: 70% Cu; 30% Zn	8522	0.385	111	3.412	88	. . .	128	144	147	147	. . .	. . .	. . .	. . .
German silver: 62% Cu; 15% Ni; 22% Zn	8618	0.394	24.9	0.733	19.2	. . .	31	40	45	48	. . .	. . .	. . .	. . .
Constantan	8922	0.410	22.7	0.612	21	. . .	22.2	26	. . .	. . .	. . .	. . .	. . .	. . .
Magnesium														
Pure	1746	1.013	171	9.708	178	171	168	163	157	. . .	. . .	. . .	. . .	. . .
Mg-Al (electrolytic): 6-8% Al; 1-2% Zn	1810	1.00	66	3.605	. . .	52	62	74	83	. . .	. . .	. . .	. . .	. . .
Molybdenum	10,220	0.251	123	4.790	138	125	118	114	111	109	106	102	99	92
Nickel														
Pure (99.9%)	8906	0.4459	90	2.266	104	93	83	73	64	59	. . .	. . .	. . .	. . .
Ni-Cr														
90% Ni; 10% Cr	8666	0.444	17	0.444	. . .	17.1	18.9	20.9	22.8	24.6	. . .	. . .	. . .	. . .
80% Ni, 20% Cr	8314	0.444	12.6	0.343	. . .	12.3	13.8	15.6	17.1	18.0	22.5	. . .	. . .	. . .
Silver														
Purest	10,524	0.2340	419	17.004	419	417	415	412	. . .	. . .	. . .	. . .	. . .	. . .
Pure (99.9%)	10,524	0.2340	407	16.563	419	410	415	374	362	360	. . .	. . .	. . .	. . .
Tin (pure)	7304	0.2265	64	3.884	74	65.9	59	57	. . .	. . .	. . .	. . .	. . .	. . .
Tungsten	19,350	0.1344	163	6.271	. . .	166	151	142	133	126	112	76	. . .	. . .
Zinc (pure)	7144	0.3843	112.2	4.106	114	112	109	106	100	93	. . .	. . .	. . .	. . .
Diamond														
Natural (Type 1a)[(a)]	3515	0.510	800	45	1500	900	600	. . .	. . .	. . .	. . .	. . .	. . .	. . .
Synthetic (polycrystalline)	3515	0.510	2000	110	6000	2200	1300	. . .	. . .	. . .	. . .	. . .	. . .	. . .
Aluminum oxide														
Sapphire[(a)(b)]	3980	0.758	40	1.326	125	46	26	18	14	12	9	8	7.5	8
Polycrystalline[(c)]	3900	0.752	30	1.02	. . .	. . .	. . .	. . .	. . .	13	. . .	. . .	6	. . .
Silicon carbide[(c)]	3200	0.670	50	2.33	. . .	. . .	. . .	40	. . .	34	30	25	. . .	. . .
Silicon nitride[(c)]	3200	0.710	30.7	1.35	. . .	. . .	. . .	27	. . .	23	20	18	. . .	. . .
Titanium carbide[(c)]	6000	0.543	55	1.69	. . .	. . .	. . .	. . .	. . .	32	. . .	. . .	28	. . .
Titanium diboride[(c)]	4500	. . .	75	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
Tungsten carbide[(c)]	15,100	0.205	102	3.3	. . .	. . .	97	92	. . .	82	. . .	. . .	. . .	. . .
Graphite[(c)]	1900	0.71	178	13.2	. . .	. . .	. . .	. . .	. . .	112	. . .	. . .	62	. . .
Nylon[(c)]	1140	1.67	0.25	0.013	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
Reinforced nylon[(c)]	1420	. . .	0.22-0.48	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
Teflon[(c)]	2200	1.05	0.24	0.010	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
Silicon oxide (glass)	2200	0.8	1.25	0.08	. . .	. . .	. . .	1.05	. . .	1.25	1.4	1.6	1.8	. . .

Source: Ref 7

(a) Materials are anisotropic and values vary with crystallographic orientation.

(b) Single-crystal synthetic material.

(c) Typical properties of bearing-quality materials. Ceramics are hot pressed or sintered. These properties are representative and depend on detailed composition and processing.

If both bodies are of the same material, the thermal contact coefficients are equal. Thus, $b_1 = b_2 = b$, and Eq 12 becomes:

$$T_f = 1.13 \cdot \left(\mu \frac{W}{L}\right) \cdot \left|\sqrt{V_1} - \sqrt{V_2}\right| \cdot (b\sqrt{w})^{-1}$$ **(Eq 14a)**

where T_f is in °C.

By utilizing Hertzian contact theory (Ref 3) to relate load and contact geometry to pressure:

$$T_f = 2.49 \cdot \mu \cdot p_H^{3/2} \cdot \left|\sqrt{V_1} - \sqrt{V_2}\right| \cdot \left(\frac{R}{E_v}\right)^{1/2} \cdot b^{-1}$$ **(Eq 14b)**

or:

$$T_f = 0.63 \cdot \mu \cdot \left(\frac{W}{L}\right)^{3/4} \cdot \left|\sqrt{V_1} - \sqrt{V_2}\right| \cdot \left(\frac{E_v}{R}\right)^{1/4} \cdot b^{-1}$$ **(Eq 14c)**

in which:

$$E_v = E/(1 - v^2)$$ **(Eq 15a)**

where E is Young's modulus (in N/m^2) and v is Poisson's ratio, and

$$R = \left[\frac{1}{R_1} \pm \frac{1}{R_2}\right]^{-1}$$ **(Eq 15b)**

where R is the equivalent radius, in meters, of convex to convex (+) or concave (-) undeformed surfaces; and p_H is the maximum Hertzian contact pressure (in N/m^2). Note that the coefficients 1.13, 2.49, and 0.63 are associated with a uniform heat flux as discussed in the section "General Contact Analysis" in this article.

Should the components have different bulk temperatures, T_{b_1} and T_{b_2}, let:

$$T_c = T_f + T_b = T_{f_1} + T_{b_1} = T_{f_2} + T_{b_2}$$ **(Eq 16)**

thus permitting, where T_c is in °C, Eq 12, 14a, 14b, or 14c to still be used for determining T_f, provided T_b in Eq 11 or 16 is replaced with:

$$T_b = T_{b_1} + (T_{b_2} - T_{b_1}) \cdot (n + 1)^{-1}$$ **(Eq 17)**

where T_b is in °C and n is determined from Eq 10b. It should be noted that the corrected heat partition factor in this case becomes:

$$\gamma_1 = \frac{n}{n+1} + \left(\frac{1}{n+1}\right)\left(T_{b_2} - T_{b_1}\right) \cdot \frac{\sqrt{\pi}}{2} \cdot \left(\frac{L}{\mu W}\right) \cdot \frac{b_1 \sqrt{w V_1}}{|V_1 - V_2|} \quad \text{(Eq 18)}$$

For the situation in which surface *i* is stationary (Pe < 2), Blok (Ref 8) suggests that the maximum flash temperature, in °C, is:

$$T_{f_i} = \frac{Q_i}{\sqrt{\pi} \cdot L \cdot k_i} \quad \text{(Eq 19)}$$

where Q_i is the rate of heat supplied to body *i.* Coupled with the moving body relationship of Eq 8, a frictional heating assessment can be made via the method discussed in the section "Circular Contact Analysis with One Body in Motion" in this article.

Circular Contact Analysis with One Body in Motion. Figure 2 illustrates a convenient model for flash temperature estimation (Ref 7), where a protuberance on body 1 forms a circular contact area of radius *a* with the flat surface of body 2. While this model is useful in representing two general surfaces forming an apparent area of contact, it is typically associated with the real area of contact formed by a pair of spherical asperities. This does not involve any loss of generality because the static contact of two rough surfaces is, to within a good degree of approximation, equivalent to the contact of a smooth surface and a rough surface with a composite roughness (Ref 9).

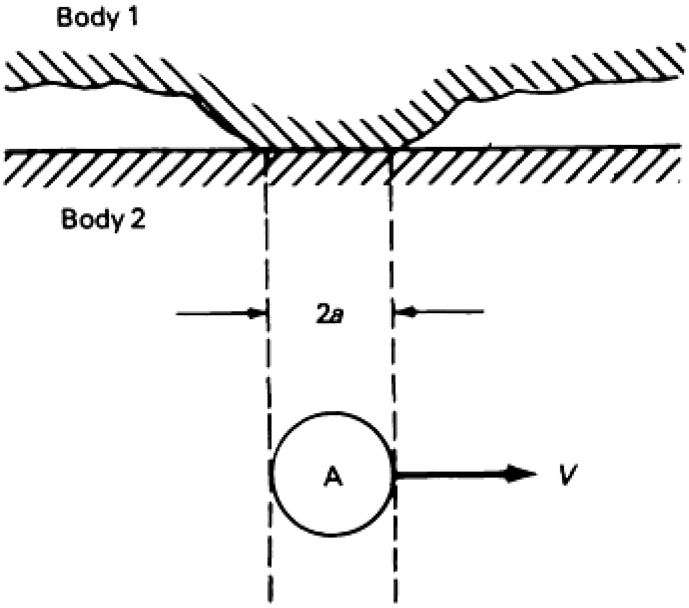


Fig. 2 Schematic showing key parameters required to estimate flash temperature using a circular contact model with a single contact area. Body 1 is stationary, while body 2 moves relative to the contact area. See text for discussion. Source: Ref 7

Body 1 is assumed to be stationary, whereas body 2 moves relative to the heat source or contact area with velocity V_2. Considering a uniformly distributed heat flux, $q = q_{av}$, over the contact width $2a$, Eq 8 can be used to establish the maximum flash temperature of the moving surface. By using a thermal-electrical resistance analogy, Holm (Ref 10) determined that the maximum flash temperature (in °C) of the stationary surface is:

$$T_{f_1} = \frac{Q_1}{\pi \cdot a \cdot k_1} \quad \textbf{(Eq 20)}$$

where Q_1 represents the rate of heat supplied to the stationary body of thermal conductivity k_1.

By equating the maximum flash temperatures of the respective surfaces in contact through Eq 8 and Eq 20, and assuming equivalent bulk temperatures with a time of contact t, equal to $2a/V_2$, the fraction of the total heat entering the stationary body 1 is:

$$\gamma_1 = \frac{k_1}{k_1 + 0.627\, k_2 (V_2 \cdot a/D_2)^{1/2}} \quad \textbf{(Eq 21a)}$$

Hence, from Eq 20, the maximum flash temperature(in °C) for use in Eq 11 is:

$$T_f = \gamma_1 \cdot \mu \cdot \frac{W \cdot V_2}{\pi \cdot a \cdot k_1} \quad \textbf{(Eq 21b)}$$

where W is the normal contact load (in N), k_i is the thermal conductivity of body i (in W/m · °C), D_2 is the thermal diffusivity of body 2 (in m^2/s), and V_2 is the velocity of surface 2 (in m/s).

Should the components have different bulk temperatures, T_{b_1} and T_{b_2}, the maximum contact temperature, T_c (in °C), can be established using Eq 16, 21a, and 21b, provided:

$$T_b = T_{b_1} + (T_{b_2} - T_{b_1}) \cdot (s + 1)^{-1} \quad \textbf{(Eq 22a)}$$

where:

$$s = 1.60 \cdot (V_2 \cdot a/D_2)^{-1/2} \cdot k_1/k_2 \quad \textbf{(Eq 22b)}$$

Although this situation utilizes Eq 21a for γ_1, the portion of heat withdrawn by surface 1 is adjusted as follows:

$$\gamma_1 = \frac{s}{s+1} + \left(\frac{1}{s+1}\right)\left(T_{b_2} - T_{b_1}\right) \cdot \frac{\pi \cdot a \cdot k_1}{\mu W V_2} \quad \textbf{(Eq 23)}$$

The analyst is reminded that should both surfaces be moving (Pe $\geq$2), the theory for the moving surface is applied to both bodies as described in the section "Line Contact Analysis with Two Bodies in Motion" in this article.

Average Flash Temperature Estimation via Heat Partitioning. Upon frictional heating, the average surface temperature over a given real area of contact, $\overline{T}_c$ (in °C) can be expressed as:

$$\overline{T}_c = \overline{T}_{f_i} + T_{b_i} \quad \textbf{(Eq 24)}$$

where $\overline{T}_{f_i}$ reflects the average surface temperature increase of body i from its bulk temperature, T_{b_i}. For a stationary body (case 1), subjected to a uniform circular heat source of radius a, Holm (Ref 10) reports that the average flash temperature (in °C) is:

$$\overline{T}_{f_1} = \frac{Q_1}{4\, a \cdot k_1} \qquad \textbf{(Eq 25)}$$

where Q_1 represents the rate of heat supplied to body 1 of thermal conductivity k_1. For a component (case 2), moving with velocity V_2 past the uniform circular heat source, the average exposure time will be $(a \cdot \pi)/(4 \cdot V_2)$ (Ref 11). Applying this parameter to Eq 8, the average flash temperature (in °C) is:

$$\overline{T}_{f_2} = \frac{Q_2}{\pi a k_2} \sqrt{\frac{D_2}{V_2 \cdot a}} \qquad \textbf{(Eq 26)}$$

where Q_2 represents the rate of heat supplied to body 2 of thermal conductivity k_2 and diffusivity D_2.

In the model of Fig. 2, surface 1 is assumed to be stationary. By assuming equivalent bulk temperatures, the average flash temperatures of the respective surfaces in circular contact may be equated. Through Eq 25 and 26, the fraction of the total heat entering the body 1 will be:

$$\gamma_1 = \frac{k_1}{k_1 + 0.785\, k_2 \,(V_2 \cdot a/D_2)^{1/2}} \qquad \textbf{(Eq 27a)}$$

Hence, from Eq 25, the average flash temperature for use in Eq 24 is:

$$\overline{T}_{f_{1,2}} = \gamma_1 \cdot \mu \cdot \frac{WV_2}{4a \cdot k_1} \qquad \textbf{(Eq 27b)}$$

Average Flash Temperature Estimation via Contact Size. Archard (Ref 12) considered special cases where the contact size is determined by either elastic or plastic deformation utilizing the model of Fig. 2. Surface 1 is assumed to be stationary, with both bodies composed of thesame material. If the deformation is plastic, the contact radius, a (in meters), is:

$$a = \left[\frac{W}{\pi \cdot p_M}\right]^{1/2} \qquad \textbf{(Eq 28a)}$$

where p_M is the flow pressure or hardness of the material (in N/m^2) and is about three times the yield strength of the material in uniaxial compression (Ref 13). If the deformation is elastic:

$$a = 0.91\left[\frac{W \cdot R}{E_v}\right]^{1/3} \qquad \textbf{(Eq 28b)}$$

where R is the undeformed radius of curvature of body 1 (in meters); and $E_v = E/(1 - v^2)$, where E_v is in N/m^2, and E is Young's modulus and v is Poisson's ratio.

Assuming both bodies are of the same material and bulk temperature, Archard suggests that at very slow speeds (Pe < 0.2), the total rate of heat flow, Q, is divided equally between both bodies; thus, $Q_1 = Q_2 = (\frac{1}{2})Q$. At very high speeds (Pe > 200), practically all the heat is supplied to the moving body 2; thus $Q_2 = Q$. Therefore, at high and low speeds, the following relationships result for an average flash temperature, $\overline{T}_f$ (in °C), given Eq 25, 26, 28a, and 28b:

- Plastic deformation, low speed (Pe < 0.2):

$$\overline{T}_f = \mu \frac{\sqrt{\pi p_M}}{8k} W^{1/2} \cdot V_2 \qquad \textbf{(Eq 29a)}$$

- Plastic deformation, high speed (Pe > 200):

$$\overline{T}_f = 0.31 \mu \frac{(\pi p_M)^{3/4}}{(k\rho c)^{1/2}} W^{1/4} \cdot V_2^{1/2} \qquad \textbf{(Eq 29b)}$$

- Elastic deformation, low speed (Pe < 0.2):

$$\overline{T}_f = 0.13\mu \left(\frac{1}{k}\right)\left(\frac{E_v}{R}\right)^{1/3} \cdot W^{2/3} \cdot V_2 \qquad \textbf{(Eq 29c)}$$

- Elastic deformation, high speed (Pe > 200):

$$\overline{T}_f = 0.36\mu \left(\frac{1}{\sqrt{k\rho c}}\right)\left(\frac{E_v}{R}\right)^{1/2} \cdot W^{1/2} \cdot V_2^{1/2} \qquad \textbf{(Eq 29d)}$$

-

By equating the stationary and moving flash temperature relationships of Table 1 for a given contact geometry, transition Péclet numbers of 2.5 and 4.0 result. This suggests that, for ease of analysis, a body of Pe < 2 may be considered "low speed" or fixed. For values in the range of 2 $\leq$Pe $\leq$200, a representative value for $\overline{T}_f$ may be obtained by taking an average of the low-speed and high-speed flash temperatures calculated using the appropriate deformation condition in Eq 29a, 29b, 29c, and 29d.

Correlation of Experimental Data with Calculated Values

Griffioen, Bair, and Winer (Ref 14) have measured surface temperatures using an infrared scanning camera. In operating from rolling to sliding, with loads resulting in Hertzian contact pressures of 2.0 GPa (290 ksi), data have been obtained under starved and lubricated conditions.

Temperature fluctuations as high as 2000 °C (3600 °F) were observed in an unlubricated contact between a silicon nitride pin and a sapphire disk that was subjected to a load of 8.9 N (0.91 kgf) with a sliding speed of 1.5 m/s (4.9 ft/s). Temperatures of this magnitude suggest that the local material undergoes extensive property changes, thus complicating any frictional heating calculations employed.

Measurements confirmed the power dependence of the flash temperature on normal contact loads as noted in the results of the section "Average Flash Temperature Estimation via Contact Size" in this article (that is, for low speeds [Pe < 0.2], the temperature rise was proportional to $W^{2/3}$ and at high speeds [Pe > 10], the temperature rise was proportional to $W^{1/2}$).

The power of dependence on sliding speed was found to be less than that in theory (0.7) versus 1.0 for Pe < 0.2 and 0.21 versus 0.5 for Pe > 10). This may have been the result of one or more of the following factors:

- In the experiment, there was lubricant surrounding the contact zone, which may cause more cooling than accounted for in theory
- Presence of a lubricant layer may cause an unaccounted-for temperature difference between the surface and lubricant in which the energy dissipation occurs

- Coefficient of friction decreased with sliding velocity whereas theory assumes it to be constant

The measured maximum temperature of the moving surfaces in the sliding-lubricated case agreed well with the calculated values. When both surfaces were moving, the measured maximum temperature correlated well with the calculated average temperature.

Factors Limiting the Accuracy of Calculations

Attempts to improve the accuracy of calculations associated with frictional heating have occurred utilizing a variety of analytical methods, with suppositions built upon the basic concepts of the section "Idealized Models of Sliding Contact" in this article. From the fundamental contributions of Jaeger (Ref 15), to the unique approximations of Ling (Ref 16, 17), Francis (Ref 18), Lim and Ashby (Ref 19), and others, sliding contact analyses have progressed, although they are limited by the complexities of the equations that result. Consequently, the generalizations that follow can be of significance when addressing alternative designs.

Effect of Transient Temperature. Using the model of Fig. 2, the contact of a stationary circular asperity (surface 1) of radius *a* creates a heat source with the moving surface of a relatively larger body (surface 2). Because the primary interest is the temperature in the immediate vicinity of the contact, the rest of the surface is assumed to be insulated.

From the governing differential equations (Ref 1), Fig. 3 was developed to show that the surface temperature behaves as a function of time, *t*. Presented in nondimensional terms of temperature (*kT/qa*), position (*r/a*), and time ($F_0^{1/2}$), the figure confirms that a steep temperature rise near the center of the source is established rapidly (for example, $F_0^{1/2}$ ~1). (Note that F_0, the Fourier modulus, is equal to $D_i \cdot t/a^2$ as discussed in the section "Heat Source Time" in this article.) At $F_0^{1/2}$ ~5, the center temperature is ~95% of the steady-state temperature, whereas the temperature at a distance *r/a* = 5 is only 50% of its steady-state value. Steady-state in the vicinity of the source is reached at $F_0^{1/2}$ ~10. Similar results were reported by Barber (Ref 20) and Greenwood (Ref 21) using asymptotic analysis.

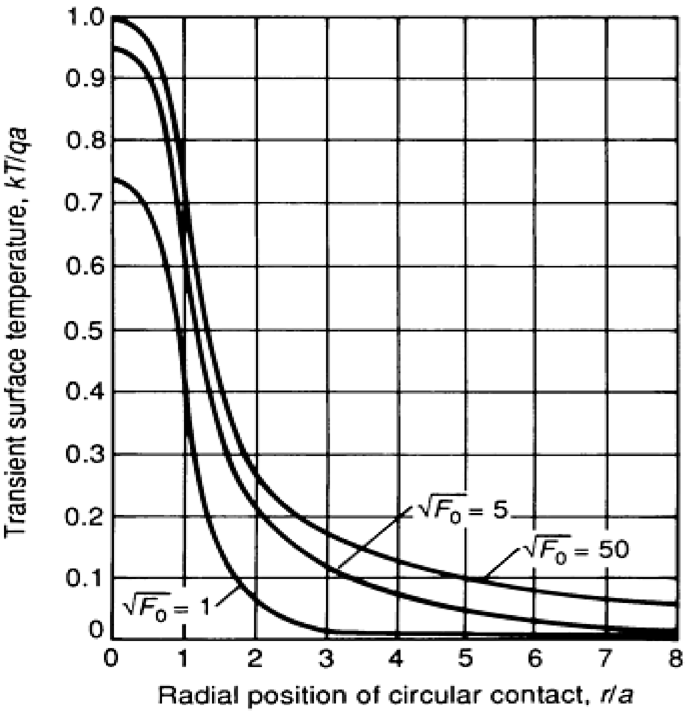


Fig. 3 Plot of surface temperature versus radial position of circular contact as a function of time. Time is

expressed in terms of Fourier modulus, F_0. Source: Ref 1

Recalling that the circular contact analysis discussed in the section "Circular Contact Analysis with One Body in Motion" in this article was based on a steady-state heating duration, results could differ if the mean effective time for the moving surface is equal to the time (in seconds) the heat source exists. That is:

$$t = (F_0 \cdot a^2)/D_2 \qquad \textbf{(Eq 30)}$$

Substituting this heat source time into Eq 8, and equating the mean temperatures of the moving and fixed surfaces, the heat partition factor γ_1, associated with the total frictional heat input to the fixed surface becomes a function of F_0. For $F_0 \sim 0$ (Ref 11):

$$\gamma_1 = \frac{k_1}{k_1 + k_2\sqrt{D_1/D_2}} \qquad \textbf{(Eq 31)}$$

Effect of Surface Layer. A surface film with different thermal and mechanical properties than those of the substrate will affect the frictional heating behavior at a sliding contact. It will not only affect the coefficient of friction, but may also affect the dissipation of thermal energy. By modeling a rotating layered cylinder, Gecim and Winer (Ref 22) closely approximated the thermal condition of the semi-infinite body (surface 2) of Fig. 2. Subjected to a "small" heat source, while incorporating a film depth δ, the body is assumed to be moving at a rate sufficient for Pe > 100.

Figure 4 was developed from the governing differential equations (Ref 22) to show how the surface temperature may behave as a function of the thermal contact coefficients of the substrate (b_s) and film (b_f). Presented in nondimensional form, the plot gives an indication of the surface temperature rise with the film, T_{f_w} to that without the film, $T_{f_{wo}}$. (Note that $T_{f_{wo}}$ is computed utilizing the models discussed in the section "Idealized Models of Sliding Contact" of this article.) The horizontal axis is associated with the product of the dimensionless film thickness and the square root of the Péclet number. Expressed as $(\delta/R_0) \cdot [(VR_0)/(2 \cdot D_f)]^{1/2}$, R_0 represents the surface radius (with film), V the surface velocity, and D_f the thermal diffusivity of the film.

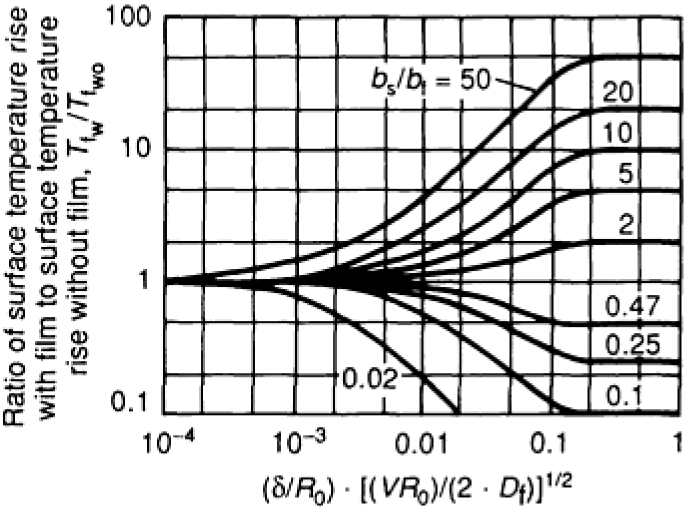


Fig. 4 Surface layer temperature graph showing effect of substrate (b_s) and film (b_f) thermal contact coefficients on surface temperature. Angular contact width, Φ_0, is 0.05 rad. Source: Ref 22

For an angular contact width Φ_0 = 0.05 rad (which is typical of a cam, gear, or rolling-element bearing contact), results suggest that for the thin-film/low-speed condition, thermal penetration goes well below the film into the substrate; thus, film thermal resistance becomes insignificant, and the temperature ratio approaches unity. For the thick-film/high-speed

case, the temperature rise is confined within the film; hence, the flash temperature is inversely proportional to the thermal contact coefficient of the film alone. As a result, the temperature ratio approaches the ratio of the thermal contact coefficients, b_s/b_f. Recall from Eq 13 that $b = (\rho \cdot c \cdot k)^{1/2}$, where ρ is the density; c is the specific heat; and k is the thermal conductivity of the appropriate material.

Limitations of Calculations. As suggested in this work and by the reviews of Blok (Ref 23) in 1970 and Kuhlmann-Wilsdorf (Ref 24) in 1987, frictional heating calculations are quite imperfect because they are typically not well defined and vary greatly in time and place. This occurs no doubt because of the lack of knowledge about many of the underlying assumptions.

A number of factors affect the presence of thermal energy from frictional resistance. Such parameters as velocity and normal load are measurable. However, material properties are constantly changing, particularly when exposed to high temperatures and heavy loads. Changes in the coefficient of friction, which enters the flash temperature computation to the first power, must be considered. Difficulties also arise when attempting to quantify the real area of contact, given its relationship to the sharing of load among contact spots that vary in size, number, and hardness.

References

1. B. Gecim and W.O. Winer, Transient Temperatures in the Vicinity of an Asperity Contact, *J. Tribol.,* Vol 107, 1985, p 333-342
2. H. Wolf, *Heat Transfer,* Harper & Row, 1983, p 193
3. S. P. Timoshenko and J.N. Goodier, *Theory of Elasticity,* 3rd ed., McGraw-Hill, 1970, p 409
4. T.F.J. Quinn and W.O. Winer, An Experimental Study of the "Hot-Spots" Occurring during the Oxidational Wear of Tool Steel on Sapphire, *J. Tribol.,* Vol 109 (No. 2), 1987, p 315-320
5. H.S. Carslaw and J.C. Jaeger, *Conduction of Heat in Solids,* 2nd ed., Oxford University Press, London, 1959, p 75
6. H. Blok, The Flash Temperature Concept, *Wear,* Vol 6, 1963, p 483-494
7. W.O. Winer and H.S. Cheng, Film Thickness, Contact Stress, and Surface Temperatures, *Wear Control Handbook,* M.B. Peterson and W.O. Winer, Ed., American Society of Mechanical Engineers, 1980, p 121-141
8. H. Blok, Theoretical Study of Temperature Rise at Surfaces of Actual Contact under Oiliness Lubricating Conditions, *Proc. General Discussion on Lubrication and Lubricants,* Vol 2, Institute of Mechanical Engineers, 1937, p 222-235
9. E. Ioannides and J.C. Kuijpers, Elastic Stress below Asperities in Lubricated Contacts, *J. Tribol.,* Vol 108, 1986, p 394-402
10. R. Holm, Calculation of the Temperature Development in a Contact Heated in the Contact Surface and Application to the Problem of the Temperature Rise in a Sliding Contact, *J. Appl. Phys.,* Vol 19, 1948, p 361-366
11. B.Y. Ting, "Thermomechanical Wear Theory," Ph.D. dissertation, Georgia Institute of Technology, 1988, p 65-73
12. J.F. Archard, The Temperature of Rubbing Surfaces, *Wear,* Vol 2, 1958/1959, p 438-455
13. D. Tabor, *The Hardness of Metals,* Oxford University Press, London, 1951
14. J.A. Griffioen, S. Bair, and W.O. Winer, Infrared Surface Temperature Measurements in a Sliding Ceramic-Ceramic Contact, *Mechanisms and Surface Distress,* Proceedings of Leeds-Lyon Symposium, D. Dowson *et al.,* Butterworths, 1985, p 238-245
15. J.C. Jaeger, Moving Sources of Heat and Temperature at Sliding Contacts, *J. Proc. R. Soc. (NSW),* Vol 76, 1942, p 203-224
16. F.F. Ling and S.L. Pu, Probable Interface Temperatures of Solids in Sliding Contact, *Wear,* Vol 7, 1964, p 23-24
17. F.F. Ling and W.M. Lai, Surface Mechanisms of Layered Media, *Solid Contact and Lubrication,* Special Publication AMD, Vol 39, American Society of Mechanical Engineers, 1980, p 27-50

18. H.A. Francis, Interfacial Temperature Distribution within a Sliding Hertzian Contact, *ASLE Trans.*, Vol 14, 1970, p 41-54
19. S.A. Lim and M.F. Ashby, Wear-Mechanism Maps, *Acta Metall.*, Vol 35 (No. 1), 1987, p 1-24
20. J.R. Barber, An Asymptotic Solution for the Short Time Transient Heat Conduction between Two Similar Contacting Bodies, *Int. J. Heat Mass Transfer*, Vol 32, 1989, p 943-949
21. J.A. Greenwood, Thermal Transient Contact Resistance, *Int. J. Heat Mass Transfer*, Vol 34 (No. 9), 1991, p 2287-2290
22. B. Gecim and W.O. Winer, Effect of a Surface Film on the Surface Temperature of a Rotating Cylinder, *J. Tribol.*, Vol 108, 1986, p 92-97
23. H. Blok, "The Postulate about the Constancy of Scoring Temperature; Interdisciplinary Approach to the Lubrication of Concentrated Contacts," SP-237, Vol 1, National Aeronautics and Space Administration, 1970, p 4.1-4.114
24. D. Kuhlmann-Wilsdorf, Demystifying Flash Temperatures: 1. Analytical Expressions Based on a Simple Model, *Mater. Sci. Eng.*, Vol 93, 1987, p 107-117

Laboratory Testing Methods for Solid Friction

K.G. Budinski, Eastman Kodak Company

Introduction

THE MEASUREMENT OF FRICTION FORCE and the calculation of the coefficient of friction are often warranted for many tribosystems. Such testing is particularly critical for brakes, clutches, and similar power- and motion-control mechanisms where the friction force must be known and controlled. If the friction characteristics of a system are somehow changed, the results can be disastrous. For example, continuous braking of an automobile on a steep incline can lead to "glazing" of the brake pad material: the surface overheats and chars, forming a material with friction characteristics different from those of the bulk, and the resulting friction coefficient may be much lower than the design coefficient of friction. Consequently, the brakes may not work at the extremes of operating conditions. Thus, manufacturers of brake and clutch facings must know the friction characteristics of their material against the counterfaces it will meet under all operating conditions. This holds true for manufacturers of many different products that experience friction, including flooring and tires.

In machine design, friction characteristics must be known for friction drives, for sizing motors, for determining work forces, and for designing any part of a machine where changes in forces between sliding members could affect operation. Friction is not a measure of wear or the tendency to wear; therefore, it does not determine service life. Frictional heating sometimes causes a machine part to fail, but this usually occurs when circumstances create abnormal friction effects. To emphasize the importance of friction in machine design, it is estimated that one-third to one-half of the world's energy production is wasted through friction (Ref 1).

This article will describe the numerous techniques used to measure friction. The objective is to provide sufficient information so that a person who has never conducted a friction test can do so. Following a brief description of the historical development of friction testing, the tests and equipment currently available for friction testing and the precautions that must be taken to ensure valid test results are discussed. The final sections of this article describe how to report data and how these data may be entered into a database.

Historical Development of Friction Testing Techniques

Research into the history of tribology suggests that Greek philosophers were familiar with the concept of friction. However, Leonardo da Vinci originated the basic precepts of friction, the basic laws, and the tests used to measure and

study friction (Ref 2). Da Vinci's sketches of friction testing apparatus (Fig. 1) were made circa 1495. He used these devices to show that friction force is dependent on the force pressing two bodies together and that it is independent of the apparent area of contact between the two bodies. Da Vinci essentially introduced the concept of a friction coefficient. He found that the ratio of the friction force (*F*) to the pressing force (*N*) was about $\frac{1}{4}$; consequently, he suggested that $F = \frac{1}{4} N$. The value of $\frac{1}{4}$ is fairly accurate for the materials that da Vinci tested (wood, stone, hemp rope, bronze, and so forth) under conditions of low speed and low normal forces.

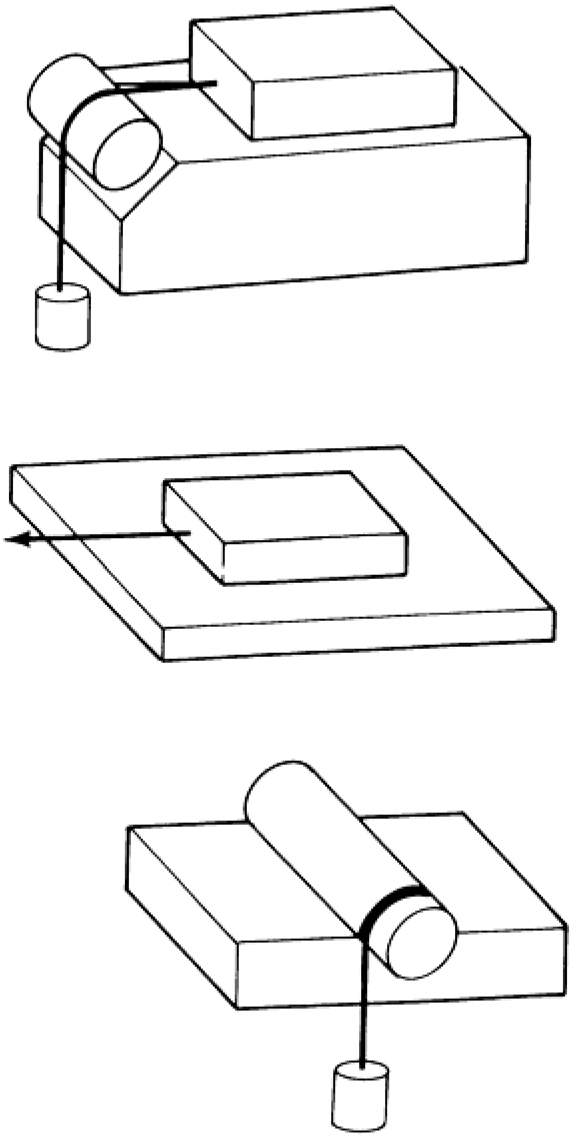


Fig. 1 Schematics of friction-measuring devices based on early work of Leonardo da Vinci

In about 1750, the mathematician Euler developed the concept of an actual friction coefficient: $\mu = F/N$. He also introduced the concept of static friction, which states that when a block is on an inclined plane, the tangent of the angle at which motion starts is the static coefficient of friction:

$$\mu_s = \tan \theta \qquad \textbf{(Eq 1)}$$

From this point on, two types of friction (static and kinetic) and two friction coefficients (the static μ_s and the kinetic μ_k) had to be considered. The static coefficient was measured in the inclined plane test, and the Kinetic was measured using

one of da Vinci's devices. Essentially the same situation exists today, except that electronic force-measuring systems may allow determination of static and kinetic friction forces on the same device in the same test.

Research in the early 19th century on the effect of lubricants on friction suggested that friction in lubricated systems depends more on the nature of the lubricant than on the nature of the materials that are in sliding contact (Ref 2). Later in that century, the Petrov equation was developed, which related the friction coefficient of a lubricated journal-bearing couple to the viscosity of the lubricant and the operational conditions of the bearing (diameter, velocity, length, and normal force). As shown in Fig. 2, this equation states that:

$$\mu = \left(\frac{4\pi r^2 l}{c}\right)\left(\frac{\eta N}{P}\right) \quad \textbf{(Eq 2)}$$

where *r* is the shaft radius, *l* is the bushing length, η is the viscosity of the lubricating fluid, *N* is the shaft velocity, *c* is the film thickness, and *P* is the normal force.

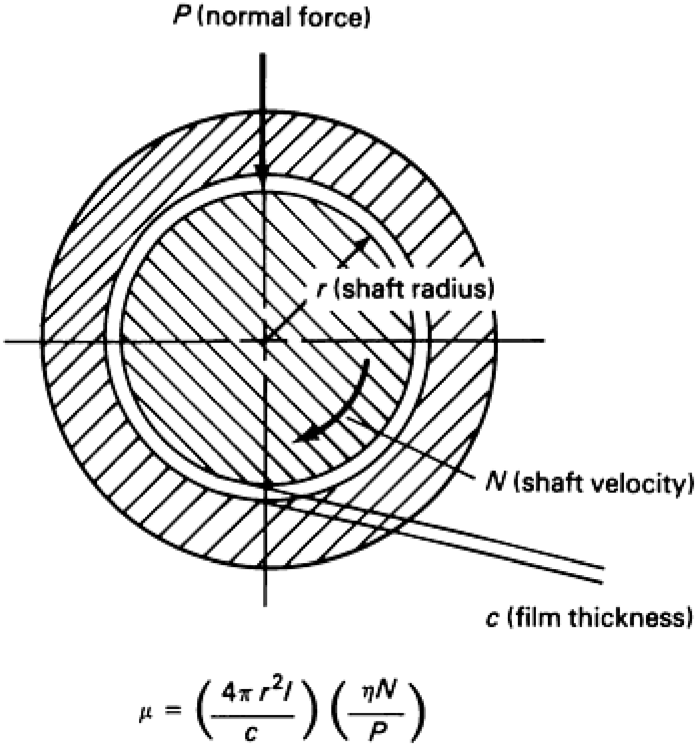


Fig. 2 Petrov equation for the coefficient of friction of a lubricated tribosystem. See text for details.

A significant amount of reseach on rolling friction was also conducted during the 19th century. It was learned that the friction of tires on surfaces is a function of the surface hardness, but a separate test for rolling friction did not emerge. By the start of the 20th century, the friction tests shown schematically in Fig. 1 were the most commonly used, and the situation remains nearly the same today.

Friction Models

There have been many models proposed for use in calculating coefficients of friction from system properties (Ref 3, 4, 5, 6, 7, 8, 9, 10, 11, 12); however, with the exception of the Petrov equation, most are not widely accepted. The Petrov equation suggests that friction in a lubricated system is a function of the system conditions and the nature of the lubricant. Using this model, which appears to work for the case of plain bearings, lubricant properties rather than friction are measured. The most accepted model for determining the coefficient of friction between solid surfaces is probably that proposed by Bowden and Tabor (Ref 5):

$$\mu_k = \frac{S}{H} \quad \textbf{(Eq 3)}$$

where μ_k is the kinetic friction coefficient, *S* is the shear strength of asperity junctions, and *H* is the bulk hardness of the softer material in the couple.

The Bowden-Tabor model is an extension of the asperity contact theory, which originated around 1930. Proponents of this theory believe that the tribological interaction between conforming solids is controlled by what happens at the point(s) where asperities on one surface make contact with asperities on the other surface. The term "asperity" usually refers to the microtopological features of a surface. Surface roughness is the most common surface property studied in conjunction with the asperity contact theory. The problem with this theory is that the equation assumes that the strength of the surface asperities is known. This information is not available and is not easy to determine experimentally. Most users of this model employ the bulk properties of the material (bulk shear strength). However, sliding conditions can obviously affect the strength of surface asperities. For example, at any sliding interface there is heating due to friction forces. This heat could affect the properties of asperities, and lower-strength materials--especially plastics--will obviously be affected (Ref 9, 11, 12). Also, it is well known that strain rate affects mechanical properties, and surface asperities may behave differently and different sliding speeds (coinciding with different strain rates).

The deformation theory of friction proposed by Bikerman (Ref 4) suggests that friction force is a function of path length and the work of deformation of surface asperities during relative motion:

$$f = \mathrm{k}wd \quad \textbf{(Eq 4)}$$

where k is a constant that depends on the mechanical properties of the two solids, *w* is the width of disturbed material, and *d* is the depth of disturbed material. This model seems intuitively correct. It simply states that the friction force is the force required to deform materials on the surface of a sliding member. This is essentially the same as the Bowden-Tabor model, but, in this case, everything is unknown. A potential user of the model knows neither the area of deformation nor the material constant.

Many other models have been proposed for calculating the coefficient of friction, but they invariably involve the use of information that is not readily available in handbooks or databases. The alternative is actual measurement of friction coefficients.

Friction Testing Techniques

Every high school physics class usually performs laboratory tests on an inclined plane to demonstrate the concept of sliding friction. The inclined plane, as demonstrated by da Vinci, remains one of the more commonly accepted techniques for measuring friction. Other techniques used to measure coefficients of friction are also modifications of concepts developed by da Vinci. Popular methods for measuring friction forces are shown in Fig. 3.

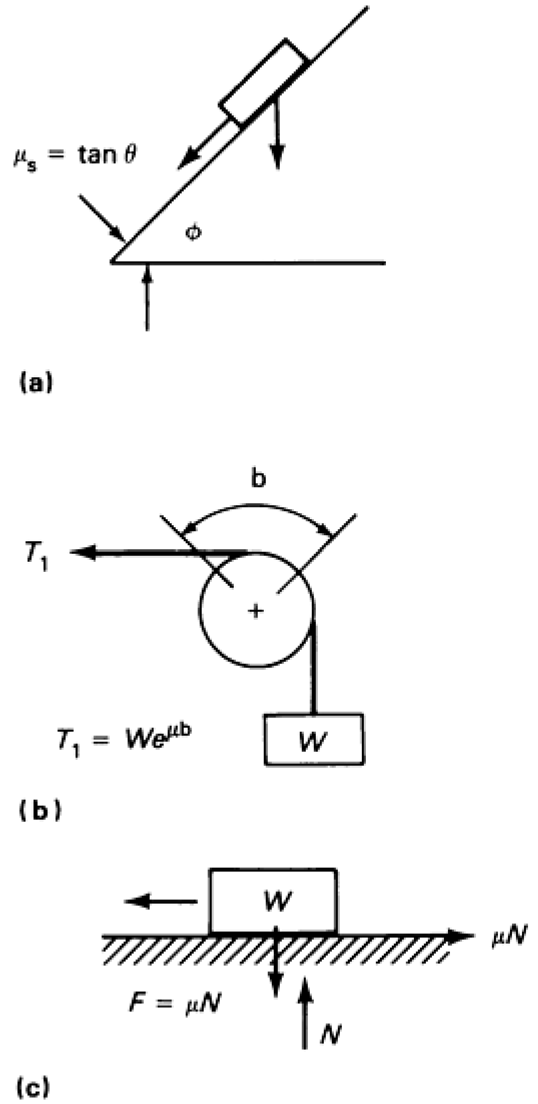


Fig. 3 Schematics of current friction-measuring concepts

In the inclined plane test (Fig. 3a), one member of the couple of interest is placed on the plane and the plane is tilted until the object starts to move. The tangent of this critical angle is the static coefficient of friction, μ= tan θ.

The problem with this technique is that it does not simulate many real-life tribosystems. Machine designers do not design systems that involve only inclined planes. Friction is a system parameter. A single material does not have a coefficient of friction; even a material pair does not necessarily have a coefficient of friction. Only when a material is mated with another material (or more than one material) and some type of sliding motion is present in the system can the coefficient of friction be measured. To be valid, a friction test should simulate the system of interest.

Figure 3(b) is classified as a "capstan test" because its principle is employed in capstans used to hold lines under tension. This friction test has utility when the system under study involves a belt, cable, rope, or web sliding over a roll or cylindrical surface. It simulates slip of plastic films on rolls during manufacturing operations, slip of belts on drive sheaves, or other practical engineering systems involving this type of tribosystem. It can be used to determine both static friction and kinetic friction, and transducers can be used to measure T_1 and W. In highspeed web handling, common practice is to place load-sensing idler rolls on either side of a system of interest (Ref 13). These rolls come in many sizes and often detect tension through strain gages mounted at the base of a cantilevered beam that supports the roll.

When using this type of system, it is important that the slip over the test roll simulate the service conditions. If a web is being transferred at 10 m/s (33 ft/s), the slip of the web on an idler (nondriven) roll will not be 10 m/s (33 ft/s), but some

fraction of this (usually less than 10%). The actual slip at a roll can be measured by instruments on the web (footage meter) and on the roll (tachometer). The difference between the two measurements yields the slip, and friction tests should be performed at this slip speed. It is also important to ensure that the surface textures on the cylinder (roll) and web duplicate the conditions in the system of interest.

The system illustrated in Fig. 3(c) can be any system where motion of one member is produced by a load cell. Pulling a block on a flat surface with a load sensor yields the friction force *F* in the basic friction equation, $F = \mu N$.

In wear studies, it is common practice to restrain a member with a load cell; the friction coefficient is calculated from this measured forced (Fig. 4). This is probably the most common method of measuring friction force. Although there are countless variations of this test, most devices use a force sensor that is essential a calibrated spring. This spring has a spring constant that determines the sensitivity of the system; there is probably some damping associated with the system. A free-body diagram of a sled friction tester is shown in Fig. 5.

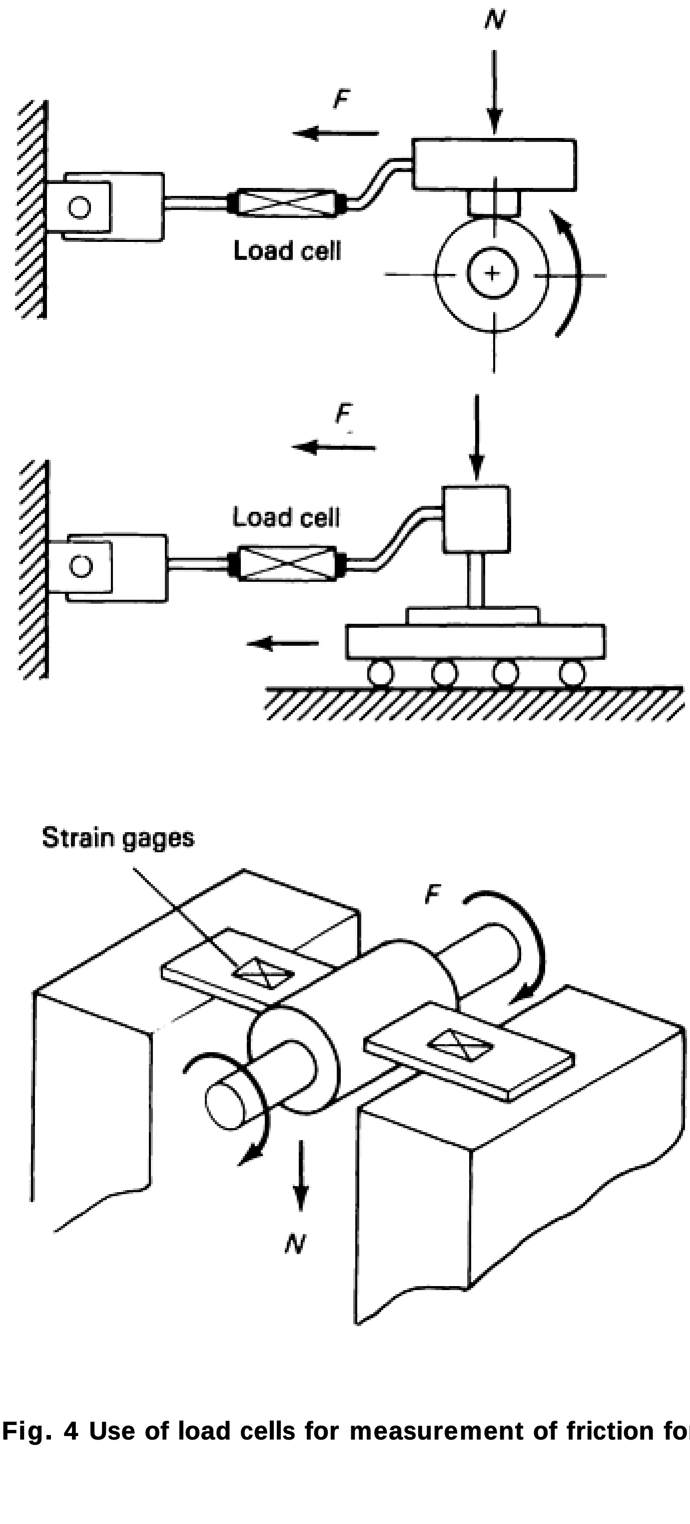


Fig. 4 Use of load cells for measurement of friction force

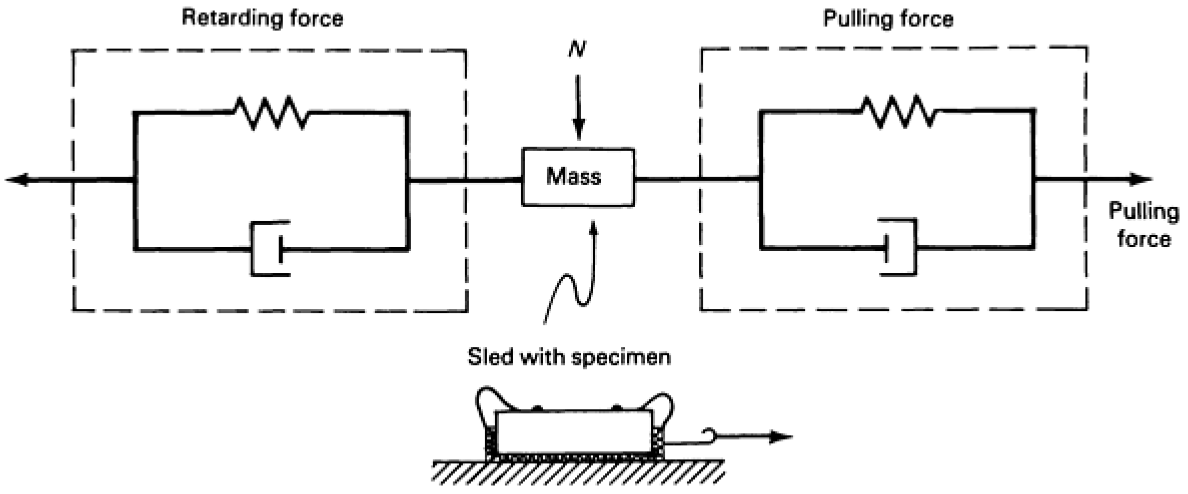


Fig. 5 Free-Body diagram of a sled friction tester

Certain system parameters can affect the results of a friction study. All that is measured is the friction force, *F,* by means of an electrical signal or a mechanical device. Both types of sensors have elastic and damping characteristics that can affect the measurement. The major concern is static friction. If the pulling system shown in Fig. 5 were replaced by a steel rod, which would have much smaller time-dependent strain behavior, the system model could be simplified as shown in Fig. 6. The measured force would still involve the elastic properties of the steel rod (spring *a*), but any time-dependent component of the friction force would come from the system under study, that is, the elasticity, *a,* and viscous behavior, *b,* of the surface contacts.

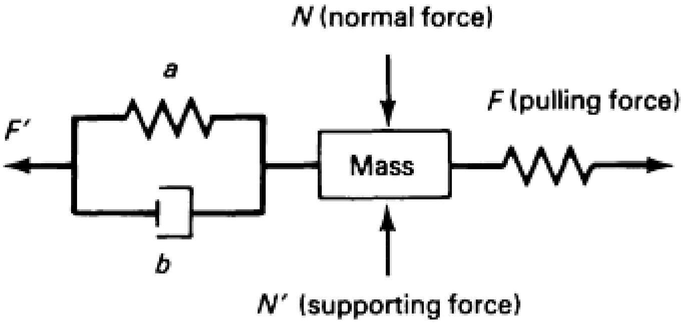


Fig. 6 Free-body diagram of a sled friction tester with a nonelastic driver

Friction Nomenclature

Terms such as "static friction," "kinetic coefficient of friction," "stick-slip behavior," and "lubricated friction," commonly used in tribology, are discussed below. Additional friction terms are defined in the "Glossary of Terms" in this Volume.

Static Friction. This term was originally defined by Amontons in 1750. In current practice, the coefficient of friction is defined in ASTM G 40 (Ref 14) as the "dimensionless ratio of the friction force between two bodies to the normal force pressing these bodies together." The static coefficient of friction is defined as the "friction coefficient corresponding to the maximum force that must be overcome to initiate macroscopic motion between two bodies." Amontons stated that in an inclined sliding-plane system, this number is equal to tan θ, the angle of inclination of the plane. This is still the case. When friction force is continuously monitored during a sliding test, a force-versus-time curve similar to the one shown in Fig. 7 is produced.

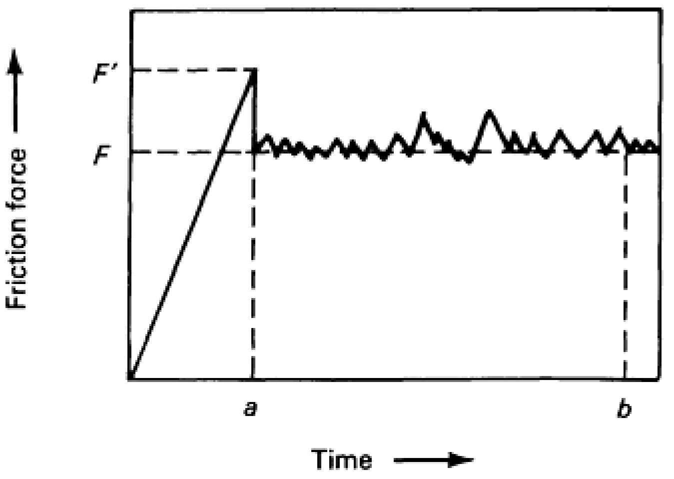


Fig. 7 Friction force recordings that are constant with time

If the force sensor detects an internal spike at the start of the test, such as *F*' in Fig. 7, this force is used to calculate the static coefficient of friction. The force *F* is used to calculate the kinetic coefficient of friction. An important consideration in a friction test is wether the internal force spike (*F*' - *F*) is due to the sliding system or is the result of elastic or viscoelastic behavior in the force-measuring system. Researchers who study static friction purposely produce motion on the movable body with a highly elastic system; for example, in a sled test, the line that pulls the sled is made of nylon or a similar low-modulus (*E*) material. If the pulling line were made of steel, there would be a significant buildup of "line stretch" before motion started. If there is a buildup of spring deflection, as in the model in Fig. 6, there will be an inertia spike caused by the basic rule of physics (force = mass × acceleration, or $F = ma$) when motion starts. The force reading *F* will sense a component from a higher acceleration *a* than would be seen if there was no buildup of spring deflection (line stretch) before motion commenced. In other words, an initial spike in a friction test can arise from spring effect or sticking of the two surfaces, or it can arise from inertial effects from the system used to produce motion:

$$F_{S_{measured}} = F_{sticking} + F_{inertia} \quad \textbf{(Eq 5)}$$

The $F_{sticking}$ between surfaces can result from reactions between surfaces, from interlocking surface features, or from other more subtle phenomena such as van der Waals forces, cohesion of surface films, and even microscopic solid static junctions between surfaces.

Static friction can be real or can be a product of the measuring system. Recommended practice for dealing with static friction in testing is to make the test model the system under study. If an initial force spike is observe, there will probably be one in the service application. If a system under study involves elastic components, a high-elasticity force-measurement system should be used.

Kinetic Coefficient of Friction. Addressing the remainder of the force-versus-time curve, the question arises as to what value should be used in measuring kinetic friction. Taking the example in Fig. 7, there are force variations that are normal in almost any sliding couple. The curve may even increase or decrease with time, as shown in Fig. 8.

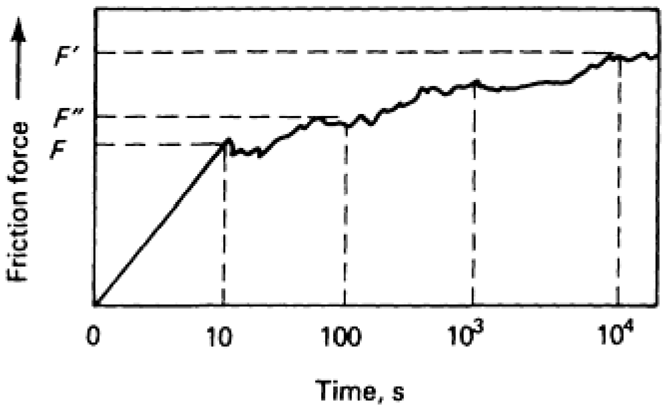


Fig. 8 Friction force recording that varies with time

ASTM G 40 (Ref 14) defines kinetic coefficient of friction as the "coefficient of friction under conditions of macroscopic relative motion between two bodies." This definition does not specify whether force *F* or *F*' or an alternate measurement shown in Fig. 8 should be used. This aspect of friction testing is left to the investigator, but it is recommended that the force number that typifies the system under study be used. If the goal of the study is determination of the friction coefficient of a part sliding down a track, and it takes 30 s for the part to do so, the force values should be averaged for that increment of time: (*F*" - *F*)/2 + *F*. If the study is concerned with friction during steady-state wear of a system after break-in and during normal running, it would be advisable to use the force readings after 10 h or even 100 h of sliding. Enough tests must be run to determine the steady-state force. If the system will wear in service, it may be desirable to measure the friction coefficient in a wear test. Wear debris between sliding members will undoubtedly affect the friction characteristics of the system.

In some standardized laboratory tests, it is customary to record the coefficient of friction after specified intervals of sliding. This provides a friction-versus-time profile and may be preferable to a fast test involving one stroke of a test device. Some investigators use computer acquisition of friction data, programming the computer to average kinetic friction over any time increment. Some test devices simply take the maximum force rendered in a test and use this as the force for static coefficient. This practice is not recommended, as the maximum force may be at the end of the test (see Fig. 8). Again, the system to use is the one that best simulates the actual system under study, but the investigator must always determine which force should be used for friction calculations.

Stick-Slip Behavior. At one time or another, everyone has encountered a squeaky door hinge that creaks when the door is opened slowly. This is likely a manifestation of stick-slip behavior. ASTM G 40 (Ref 14) defines stick-slip as "a relaxation oscillation usually associated with decrease in coefficient of friction as the relative velocity increases." In more descriptive terms, stick-slip is the jerky motion that sometimes results when one object is slid on another. If a mass is being pulled by a line, the mass will stick, then rapidly move, then restick, and so on. Under such conditions, a friction-versus-time recording may resemble that shown in Fig. 9.

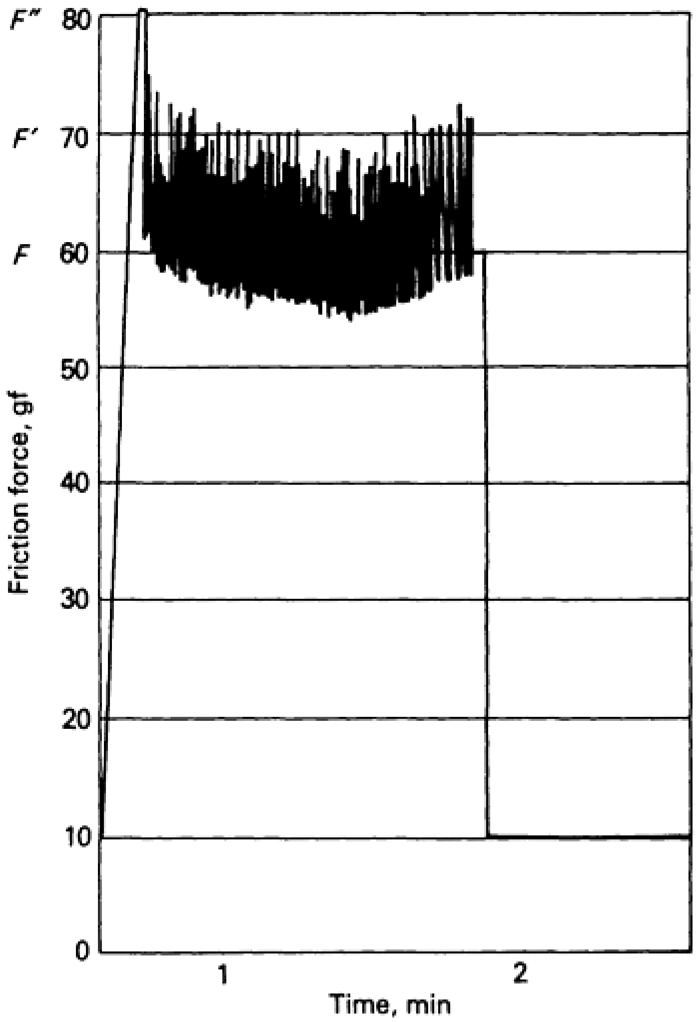


Fig. 9 Friction force recording for a system displaying stick-slip behavior

There may or may not be a friction force spike to suggest a static coefficient of friction. Stick-slip behavior in a sliding system is indicated by an obvious jerky motion, which can be recorded in a friction force versus time trace. Noise in sliding systems often results from stick-slip behavior.

One way of dealing with the coefficient of friction in a system that exhibits stick-slip behavior is to calculate the minimum and maximum coefficients by using *F* and *F*' in Fig. 9 and then noting when stick-slip occurs. In Fig. 9, *F*" could be used to calculate the static coefficient of friction. Some investigators simply ignore the numbers and state "stick-slip behavior," meaning that the kinetic coefficient of friction is not particularly meaningful for the system under study because the system produces erratic motion under the test conditions.

Sometimes stick-slip behavior results from the testing system. As discussed previously, a high-stiffness-movement system attached to the sliding member may eliminate this phenomenon. In general, if stick-slip occurs in a system under study, the test should include means of eliminating this type of behavior.

There is disagreement over the mechanism of stick-slip behavior. Some researchers believe that it occurs when the kinetic coefficient of friction is higher than the static coefficient of friction. Some say it occurs when the friction-versus-velocity plot has a negative slope (Ref 15). This is what friction recordings apparently indicate about the process, but this is not really the mechanism. Intuitively, the mechanism is probably the spring behavior of the system shown in the model in Fig. 5.

The contacts between two bodies exhibit elastic behavior, and there may be some viscoelastic behavior depending on the materials involved. In laboratory sled tests, it appears that stick-slip occurs when a sufficiently high spring force builds up

in the nylon pulling line. The specimen then moves rapidly to unload the spring (pulling line); it comes to rest when it reaches the spring deflection (line stretch), and the process begins again. What causes a material to do this is a function of the nature of the contacting materials, including adsorbed films, surface topography, the chemical nature of the surfaces, conformability, and contact stress.

Lubricated Friction. If sliding members are fully separated by a film of lubricant (oil, water, gas, and so forth), the friction coefficient of the system is essentially the friction coefficient of the fluid. When boundary lubrication exists, the surfaces are partially in contact, and the coefficient of friction of the system is a function of the combined properties of the fluid and surfaces. When dry-film (solid) lubricants are used, the friction coefficient of the system may be related to the properties of the dry-film lubricant slipping on itself or to those of an uncoated solid on the lubricant film. Lubrication with graphite, molybdenum disulfide, or fluorocarbons is an example of this (see the article "Solid Lubricants" in this Volume).

With fluid separation, the friction coefficients can be several orders of magnitude lower than for systems that involve solid contact. The Stribeck-Hersey curve shown in Fig. 10 allows calculation of the coefficient of friction from lubricant properties ($\eta N/P$), where η is the viscosity, N is the velocity, and P is the pressure (load). Empirical measurement of friction in film-separated systems is complicated by the need to maintain the fluid separation. With oils and greases, it usually involves testing at high velocities. If circumstances allow, the use of a friction coefficient calculated from the fluid parameters is infinitely easier.

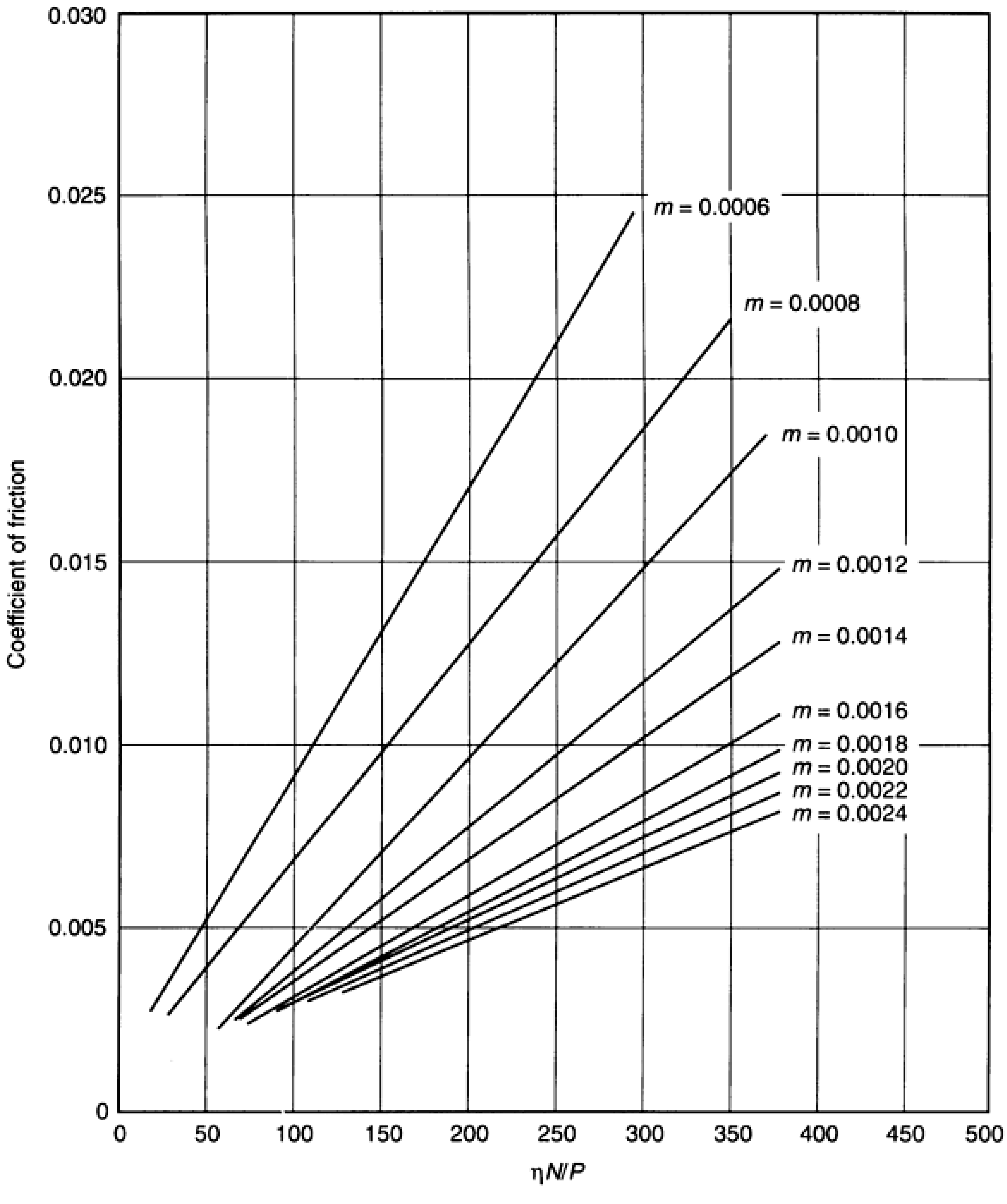


Fig. 10 Example of a Stribeck-Hersey curve for determining the friction coefficient from lubricant properties

Standard Friction Tests

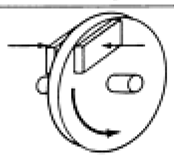
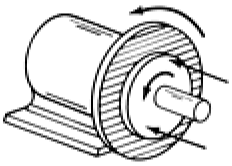
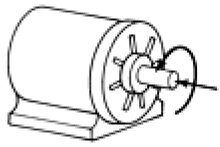
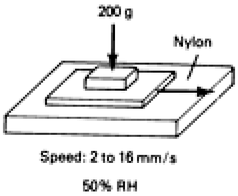
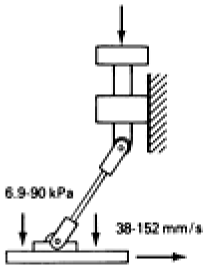
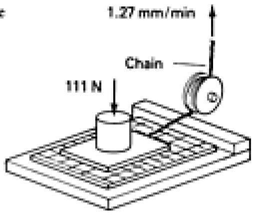
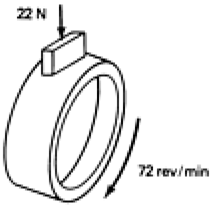
A number of organizations have developed standard tests for measuring coefficient of friction. These tests vary greatly in type and purpose.

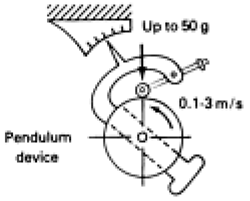
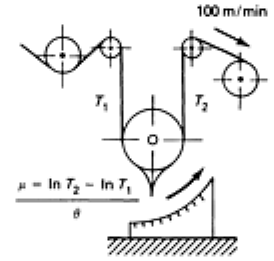
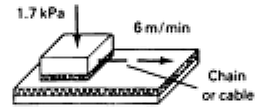
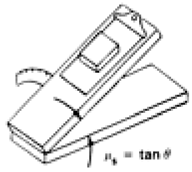
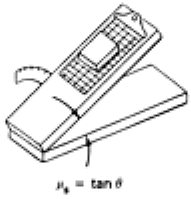
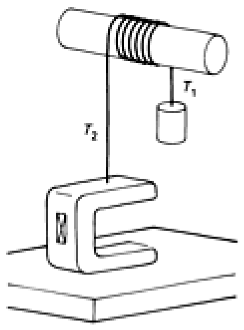
ASTM Standards and Specifications. Table 1 lists tests that have been standardized by ASTM. Some are directed toward a particular application or material, while others are for general evaluation of materials. The B-9 Committee standards deal with the types of composite friction materials used for clutches and brakes. The C-5 Committee standard is essentially a checklist of considerations for conducting coefficient of friction tests on carbon/graphite types of materials.

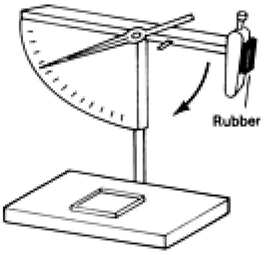
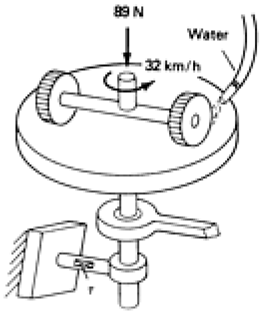
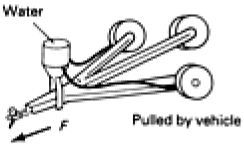
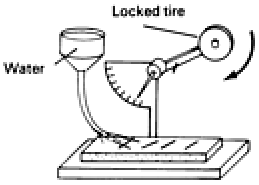
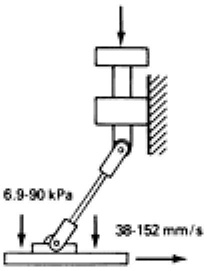
These considerations apply to most friction studies, and this standard is a suitable reference for any novice in the field of friction.

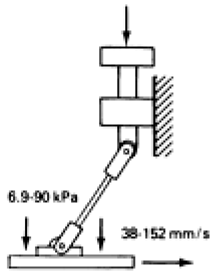
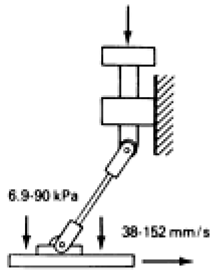
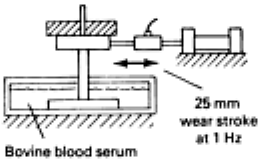
Table 1 Selected solid friction tests per ASTM standards

Note: Used with permission of the G-2 Committee on Wear and Erosion

Standard (committee)	Title	Material couple (measured parameters)	Test configuration
B 460 (B-9 on metal powders)	Dynamic Coefficient of Friction and Wear of Sintered Metal Friction Materials under Dry Conditions	Friction materials versus metal (μ_k versus temperature)	
B 461 (B-9 on metal powders)	Frictional Characteristics of Sintered Metal Friction Materials Run in Lubricants	Friction materials versus metal (μ_k versus number of engagements) (μ_k versus velocity)	
B 526 (B-9 metal powders)	Coefficient of Friction and Wear of Sintered Metal	Friction materials versus gray cast iron (μ_s and μ_k)	
C 808 (C-5 on carbon/graphite)	Reporting Friction and Wear Test Results of Manufactured Carbon and Graphite Bearing and Seal Materials	Carbon versus other materials (μ_s and μ_k)	Any apparatus
D 1894 (D-20 on plastics)	Static and Kinetic Coefficients of Friction of Plastic Films and Sheeting	Plastic film versus stiff or other solids (μ_s and μ_k)	
D 2047 (D-21 on polishes)	Static Coefficient of Friction of Polish Coated Floor Surfaces as Measured by the James Machine	Flooring materials versus shoe heels and soles (μ_s and μ_k)	
D 2394 (D-7 on wood)	Simulated Service Testing of Wood and Wood-Base Finish Flooring	Wood and wood-base flooring versus sole leather (μ_s and μ_k)	
D 2714 (D-2 on lubricants)	Calibration and Operation of Alpha Model LFW-1 Friction and Wear Testing Machine	Steel ring versus steel block (lubricated with standard oil) (μ_k)	

Standard (committee)	Title	Material couple (measured parameters)	Test configuration
D 3028 (D-20 on plastics)	Kinetic Coefficient of Friction of Plastic Solids and Sheeting	Plastic sheets or solids versus other solids (μ_s and μ_k)	
D 3108 (D-13 on textiles)	Coefficient of Friction, Yarn to Solid Material	Textile yarns versus solids (μ_k)	
D 3247 (D-6 on paper)	Coefficient of Static Friction of Corrugated and Solid Fiberboard	Self-mated cardboard (μ_s)	
D 3248 (D-6 on paper)	Coefficient of Static Friction of Corrugated and Solid Fiberboard (inclined plane method)	Self-mated cardboard (μ_s)	
D 3334 (D-13 on textiles)	Testing of Fabrics Woven from Polyolefin Monofilaments	Self-mated woven fabric (μ_s)	
D 3412 (D-13 on textiles)	Coefficient of Friction, Yarn to Yarn	Continuous filament and spun yarns self-mated (μ_s and μ_k)	 $\mu = [\ln(T_2/T_1)]/\theta$ where: μ = coefficient of friction θ = wrap angle, rad T_1 = input tension, gf (or mN) T_2 = output tension, gf (or mN)

Standard (committee)	Title	Material couple (measured parameters)	Test configuration
D 4103 (D-21 on polishes)	Preparation of Substrate Surfaces for Coefficient of Friction Testing	Vinyl and wood tiles (preparation only)	Any apparatus
E 303 (E-17 on traveled surfaces)	Measuring Surface Frictional Properties using the British Pendulum Tester	Rubber versus pavement (BPN, British Pendulum Number)	
E 510 (E-17 on traveled surfaces)	Determining Pavement Surface Frictional and Polishing Characteristics using a Small Torque Device	Rubber versus pavement (TN, torque number)	
E 670 (E-17 on traveled surfaces)	Side Force Friction on Paved Surfaces using the Mu-Meter	Tires versus pavement (μ) ($F_{dry} - F_{wet}$)	
E 707 (E-17 on traveled surfaces)	Skid Resistance of Paved Surfaces using the North Carolina State University Variable Speed Friction Tester	Rubber tire versus pavement (variable speed number)	
F 489 (F-13 on footwear)	Rating of Static Coefficient of Shoe Sole and Heel Materials as Measured by the James Machine	Leather and rubber sole and heel material versus walking surfaces (μ_s)	

Standard (committee)	Title	Material couple (measured parameters)	Test configuration
F 609	Test Method for Static Slip Resistance of Footwear, Sole, Heel or Related Materials by Horizontal Pull Slipmeter (HPS)	Footwear materials versus walking surfaces (μ_s)	
F 695-81 (F-13 on footwear)	Evaluation of Test Data Obtained by Using the Horizontal Slipmeter or the James Machine for Measurement of Static Slip Resistance of Footwear, Sole, Heel, or Related Materials	Footwear materials versus walking surfaces (reliable ranking of footwear for slip resistance) (μ_k)	
F 732 (F-4 on medical and surgical materials)	Reciprocating Pin-on-Flat Evaluation of Friction and Wear Properties of Polymeric Materials for Use in Total Joint Prostheses	Materials for human joints (μ_k)	

Note: Used with permission of the G-2 Committee on Wear and Erosion

The D-20 Committee friction test for plastic film versus other solids is a general type of test. A plastic film is placed on a platen, and then a sled, faced with a material of interest, is pulled on the surface. Although not covered in the standard, this same test works well for coated metal, sheet metal, and paper. The test uses a nylon pull line; thus, the previously mentioned system-elasticity considerations apply.

The D-21 Committee standard is very specific to floor finishes. It uses an apparatus (the James machine) that simulates the friction contact between shoe heels and soles and flooring. This same device is used for the friction tests developed by the F-13 Committee on footwear. The D-7 Committee standard on flooring addresses the friction coefficient of flooring and sole leather. All of these tests use a sample pulled by a chain. This reduces system elasticity and stick-slip behavior.

The D-2 Committee test on the block-on-ring machine is really a calibration test for the machine and is not intended for comparison of materials or lubricants. However, the apparatus described in the standard is widely used for wear testing. The test specimen is held by a friction-force transducer. ASTM G 77 describes the use of this machine for wear testing, and this standard can be used to conduct sliding wear tests on a wide variety of materials. The coefficient of friction measured in a wear test on this machine will be for worn surfaces. The lowest normal force that can ordinarily be applied in this machine is 45 N (10 lbf). The normal force is sufficient to generate damage in most materials. The only way that friction can be measured on unworn surfaces on this machine is to minimize revolutions of the counterface ring.

The rig illustrated for use in the D 3028 test standard is commonly called a "frictionometer." One test specimen is a 100 mm (4 in.) diam by 6.4 mm (0.25 in.) thick disk and the other is a 2.5 mm (0.1 in.) wide disk with an outside diameter of 20 mm (0.8 in.). The larger disk can rotate up to 200 rev/min. The smaller nonrotating disk is spring-loaded against the larger disk. The spring force is variable, and the spring provides the normal force between the two disks. The test is normally run in an accelerating and decelerating mode. The rotational speed of the larger disk is increased up to the desired terminal velocity; the speed is then slowly reduced to zero. As the friction force changes, the position of the nonrotating disk changes. At the start of a test, it may be at the 12:00 position. When the friction increases, it may move to the 1:00 position, then to the 2:00 position, and so forth. This position change is sensed by the location of a pointer on a protractor-type scale, and a coefficient of friction is determined from the number that is indicated on the scale. Friction force is measured mechanically--essentially by the rotation of a pendulum weight.

The frictionometer is easy to use, is capable of sliding speeds up to 6 m/s (20 ft/s), and can be adapted to measure the friction of a film versus the nonrotating disk. This device is especially useful for measuring the friction between plastics and metals under light loads and high velocities. This device is somewhat confusing in that the friction coefficients measured during increasing velocity often do not match the friction coefficients measured during the decreasing velocity

portion of the test. This may be due to some mechanical attribute of the machine, or it may be what actually happens at the surface.

The D 3108 test for yarn versus "solid materials" uses the capstan formula for measuring friction. If there is zero friction between the yarn and the sheave with the pointer on it (Table 1), the yarn tensions T_1 and T_2 will be the same. As the friction increases, the pointer sheave will rotate. The degrees of rotation are measured, and this number is used to calculate the coefficient of friction. This device is essentially dedicated to yarns and similar materials sliding over a curved surface at relatively high speeds (100 m/min, or 325 ft/min).

The D-6 Committee tests (D 3247 and D 3248) are sled and inclined plane tests, respectively, using classic measurements and equipment. These standards apply to self-mated cardboard. The D 3334 test is a similar classic inclined plane test for measuring static coefficient of friction. It applies to self-mated woven polyolefins, but could be used for other fabrics.

The D 3412 test, which is an extension of the D 3108 test, allows testing of self-mated yarns. The yarn is wrapped tightly on a cylinder. A single strand of yarn is then wrapped 180° around the yarn-wrapped cylinder; one end is attached to a load cell, the other to a dead-weight mass. The yarn-covered cylinder is rotated, and the friction coefficient is calculated using the capstan formula.

Specification D 4103 is not a friction test, but rather a specification for preparing vinyl and wood surfaces for friction testing. This is the same procedure used to prepare polish-coated (waxed) samples of wood and vinyl for the D 2047 "walking" machine friction test.

The E 303 test measures the friction of rubber versus concrete and similar paving materials. A block of rubber is attached to a pendulum similar to those used in impact tests. There is a preset amount of interference between the rubber and the sticky counterface. The pendulum is released; the arm swings, strikes the counterface, and the follow-through swing after the rubber strike is measured. The higher the friction, the more energy adsorbed and the shorter the back-swing after impact.

The E 510 test measures the effect of pavement finishes on the traction of tires. The desired finish is put on a sample of concrete, and two "go-cart"-size rubber tires are driven on it. The concrete disk is kept stationary by a torque-measuring arm. The end result is a number that relates to the coefficient of friction between the tires and the pavement.

The E 670 test is not a laboratory test, but instead requires pulling a trailer-type device fitted with three full-size tires. Two of the tires are at an angle different from the direction of tow. As a result, the tires skid when towed. Water is also directed at the pavement/skidding tire interface. The coefficient of friction between the pavement and the skidding tires is measured by the force that the trailer exerts on the trailer hitch of the pull vehicle. This procedure allows fullsize tires to be tested on a wide variety of pavement surfaces.

An alternate test of the E-17 Committee on traveled surfaces is a pendulum (E 707) that calls for water to be applied to the pavement sample and a small rubber tire attached to the pendulum. The test result is a variable speed number, but this number is a real measure of the friction between the tire and the pavement.

Specifications F 489, F 609, and F 695 are similar to the tests used for floor finishes; however, their purpose is to evaluate shoe soles and heels. The tests were developed by the F-13 Committee on footwear and use the James machine, which simulates walking.

Test F 732, developed by the F-4 Committee on medical and surgical materials, uses a reciprocating flat-on-flat material couple and is intended to simulate motion in human joints. The test is used to evaluate candidate couples for prosthetic devices. It is run submerged in bovine blood to simulate body conditions. The usual couples are plastic-to-metal, but many other systems have been tested.

Non-ASTM Standards. Other organizations such as the American National Standards Institute (ANSI) have also developed friction test standards. If a material under study is commonly under the jurisdiction of a particular standards organization, that organization's applicable standards should be used. For example, the photographic film industry frequently adopts ANSI and International Organization for Standardization (ISO) standards. There is an ANSI test for the static coefficient of a film in contact with a metal surface. Whenever possible, standard tests should be used. They have been agreed upon by peer groups and subsequently approved by representatives from many industries. Using such standardized tests increases the likelihood of producing valid results when compared with in-house tests.

Performing a Valid Test

The procedural considerations that should be addressed to ensure that valid data are derived from a friction test will be discussed in this section. Investigators often run a friction test that is convenient to use but that has no resemblance to the system under study. Friction is a system property, and systems must be modeled carefully (same materials, same treatments, and so forth). Other factors that can affect the validity of a friction test include sample preparation, method of friction measurement, and interpretation of data.

System Modeling. Essentially three types of geometries prevail in systems where friction is a concern. These include:

- Flat-on-flat
- Conforming surfaces (belt on a drum, plain bearings, and so forth)
- Hertzian contact

These systems are illustrated in Fig. 11. It could be argued that when wear occurs, all surfaces become conforming. This is true for all systems but those where the triboelements remain Hertzian in contact. For example, railroad wheels wear significantly in service, but they still contact the track in a Hertzian manner by line or point contact. With regard to wear of the sliding surfaces, this may be the most important aspect of modeling a tribosystem. If the system under study will wear, and it is the desired goal of the study to determine the friction characteristics under conditions of steady-state wear, friction should be measured in a wear test. If the system of concern will not change with time, a friction test that does not involve wear can be conducted. Figure 12 shows how wear can affect friction forces. Figure 12(a) illustrates the friction force variation with time for a system that does not experience any wear or changes in behavior when wear occurs. If, for example, paper is conveyed over a fixed roller with a smooth finish, it is highly likely that the friction force will remain the same as the roller wears by polishing abrasion, and the friction force will be constant with time, as will the nature of the tribosystem. The roller constantly wears, but the surface is always polished. The wear debris is carried away with the product, and nothing in the system changes. This statement must be qualified by noting that the environment (humidity, temperature, and so forth) may cause friction to change with time. Environmental effects are discussed later in this article.

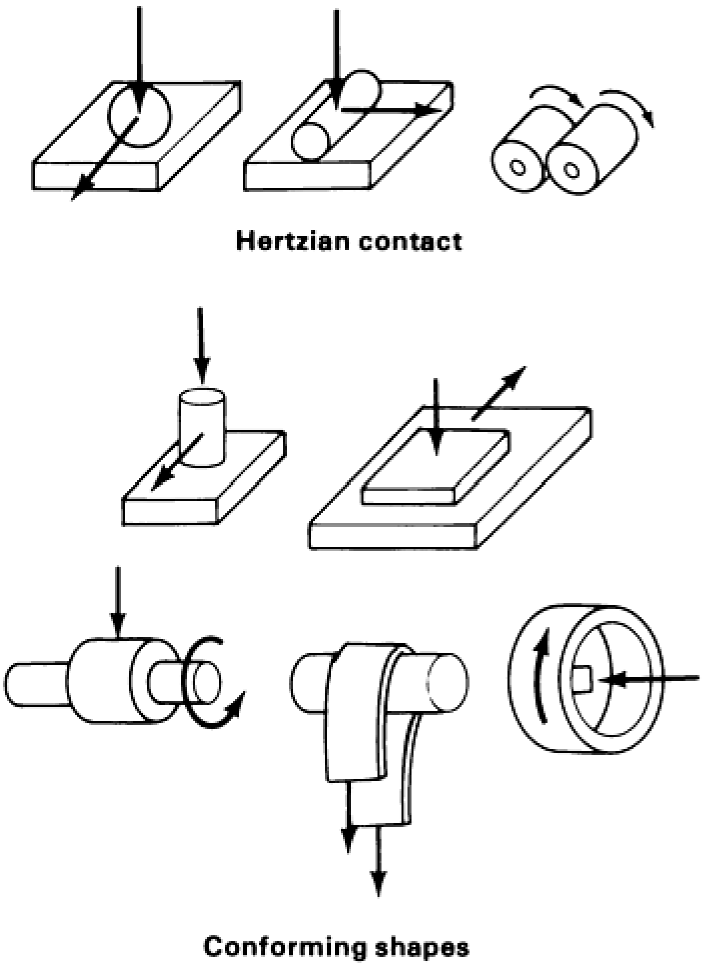


Fig. 11 Friction test specimen configuration options

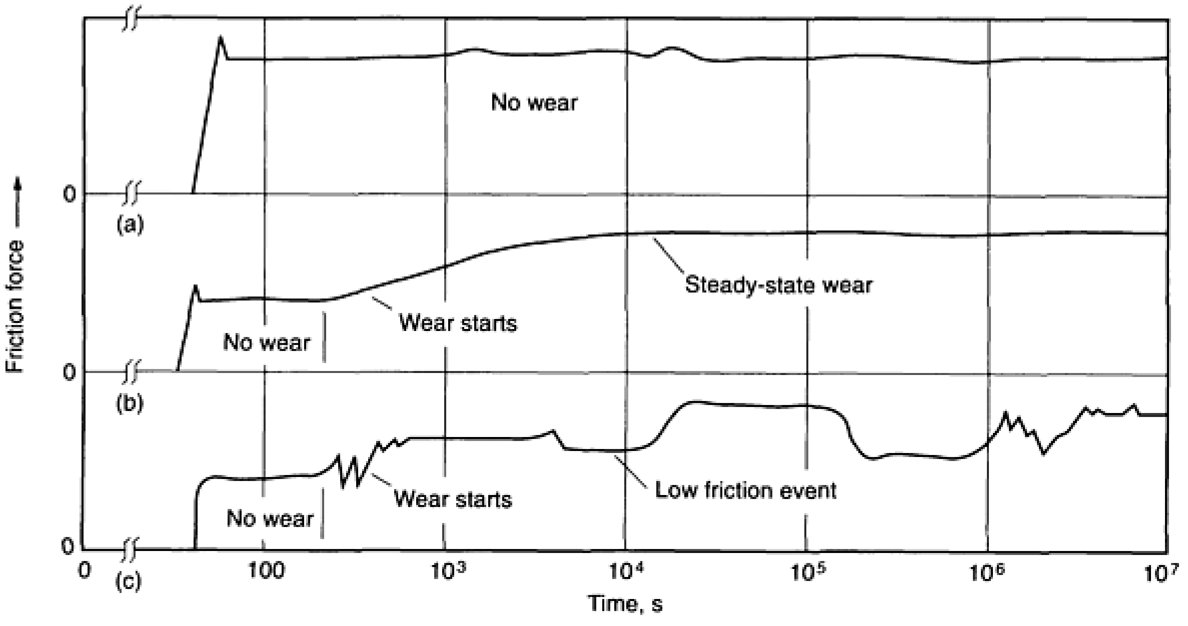


Fig. 12 Effect of system wear on friction force. (a) System that does not experience any wear or changes in behavior when wear occurs. (b) System where friction force increases with time until reaching a steady-state condition. (c) System where friction force varies with each event in the wear process

Figure 12(b) illustrates a tribosystem where the friction force increases with time and then reaches steady state. This type of behavior might occur in a system experiencing severe wear. Friction is low when both members have their original surfaces. When the original surfaces start to wear, friction increases, and from that point on, the worn surfaces remain the same. It is comparable to a machining operation in that material is constantly being removed, but the cutting forces remain constant.

Figure 12(c) illustrates a system where the friction force varies with events in the wear process. This type of behavior often occurs in a tribosystem where wear debris remains in the system. Depending on how it separates the surfaces, wear debris can have a lubricating or nonlubricating effect on the system. The friction will be constant when the triboelements have their original surfaces; when they wear, there will be a friction event followed by another when both surfaces become coated and are separated by wear debris. Because wear is a statistical process, some other event may take place to remove the separating wear debris, and still another cycle or friction event will occur. This phenomenon is seen in many machines. The machine may run quietly for a year, but then a squeal occurs that lasts for two weeks and then stops. The noise may come back in two months, or it may never return. It is quite possible that significant fluctuations in friction force with time are the result of environmental events. For example, it is well documented that the friction characteristics of photographic films are significantly affected by relative humidity. Photographic emulsions are gelatin based. Many gelatins have mechanical properties that are directly proportional to their water content. Mechanical properties vary with humidity, and these property changes in turn affect friction characteristics. Temperature has a similar influence on many materials, particularly plastics.

The illustrations shown in Fig. 12 are presented in order to emphasize that it is important to closely model the system under consideration. The test should simulate the mode of contact and the environment, and the length of time that the system is tested should be typical for the intended service. If new materials are being screened for friction characteristics, short-term and long-term tests should be conducted to explore friction events that change with time. If a service problem is being studied, the test should simulate the important aspects of the service, the environment, the type of contact, and the materials. It is not necessary to exactly duplicate the sliding speed and load unless they are likely to change some properties of the test materials (temperature, chemical environment, and so forth).

Material Documentation. Problems concerning repeatability of a friction test with time can be related to subtle or major changes in the nature of one of the tribocomponents. For this reason, it is important to provide adequate documentation of the materials in a tribosystem where friction measurement is the goal. As shown in Table 2, the necessary documentation varies with the type of material (metal, plastic, ceramic, or composite).

Table 2 Material parameters that should be documented to ensure repeatability when testing tribosystems

Metals
Chemical composition
Manufacturer's name and address
Manufacturing process [wrought (cold finished or hot finished), cast, extruded, etc.]
Generic designation
Trade name (if applicable)
ASTM or similar designation
Condition of heat treatment
Surface finish
Surface treatments
Grain direction
Microstructure
Plastics
Polymer species (If a blend, state ratio of blended polymers)
Fillers, lubricants, and so forth (percentage)
Manufacturing process (injection molded, extruded, cast)
Generic name
Applicable specifications (MIL, ASTM, and so on)
Trade name and manufacturer's number
Manufacturer's name and address
Process used to generate wear surfaces, orientation of wear surfaces to original shape (wear surfaces flatwise on as-molded flatwise plate, wear surfaces are end grain in extruded shape, and so on)
Lubricants on the surface or in the material
Surface texture of wear surfaces (R_a, lay, waviness)

Heat treatments applied
Composites
Detailed description of composite components (filament winding details, number of plys, components in laminate)
Generic name and applicable specifications on composite components
Trade names of components
Manufacturer's name and address
Method of manufacture
How wear surfaces were generated
Orientation of wear surfaces to reinforcement
Heat treatments
Surface texture (R_a, lay, waviness)
Surface treatments, lubricants, mold releases, and so forth
Ceramics
Basic type and composition
Percent theoretical density
Manufacturer's name and address
Trade name
Manufacturing process (HIP, vacuum sintering, chemical vapor deposition, and so on)
Method used to generate wear surfaces (grinding, polishing, and so on)
Texture of wear surfaces (R_a, lay, waviness)
Heat treatment
Surface treatments
Grain size
Microstructure
Morphology (degree of crystallinity, phases present, crystal structure)
Lubricants/impregnants

More items undoubtedly could be added to Table 2, but the point to be made is that all of the variables that have been mentioned thus far could have an effect on sliding friction. If the surfaces are lubricated, the tester must add related details of the type of lubricant and its distribution on the sliding surfaces. It is imperative that documentation on all members in the tribosystem be included. It is quite easy to cite examples where each of the factors listed in Table 2 has affected friction results in laboratory tests; more importantly, these factors influence friction properties in service. Heat treatment, structure, and so forth can affect mechanical properties; surface films affect sliding forces; surface texture affects early friction; a cast polyester will have different mechanical properties than a tentered and drafted polyester web; a hot isostatically pressed silicon carbide will have differentmechanical properties than pressed and sintered material. The tester should be meticulous about material documentation.

Surface Condition. Few tribologists will argue with the statement that friction can be affected by surface condition. If the loads are light enough, only the outer one or two monolayers on the surface may take part in sliding interactions. In lubricated systems, surface roughness plays a role in determining whether the surfaces are separated. Full-film separation will not occur unless the film thickness is greater than the surface roughness. Even with hydrodynamic lubrication, there are some investigators who claim that the system friction is affected by the surface roughness and lay (Fig. 13). In the normal range of roughnesses on rolling-element bearings and plain metal bearings (0.1 to 1 μm R_a), it is likely that surface features do not significantly affect lubricated sliding. In nonlubricated sliding, the effect of roughness can be negligible to significant, depending on the system (Ref 16). With normal machined surfaces with a roughness less than about 1.5 μm, the effect will be slight. Roughness becomes a meaningful factor when the roughness interactions result in mechanical locking effects. This usually happens only with very rough surfaces. Sometimes, very smooth surfaces cause adhesion--for example, gage blocks or plastic webs on smooth surfaces. Conversely, in some applications such as high-speed sliding of yarns and filaments over stationary guide pins, the friction coefficient is low when the pin surfaces are rough (0.5 to 2 μm R_a) and becomes very high when the pins become polished from wear.

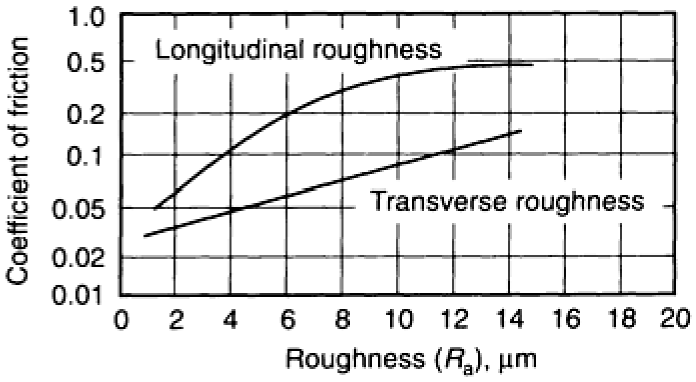


Fig. 13 Effect of surface roughness on the friction coefficient of a lubricated sliding system. Adapted from Ref 16

A valid wear test considers the effect of surface texture. For many systems, the effect is not significant. If, however, surface texture is considered a variable for a given wear study, a few quick tests using the high and low of the expected range of surface texture will determine whether this is a parameter to be pursued in a friction study.

Surface films are important in most tribosystems, especially unlubricated systems. The recommended practice in friction testing is to leave the films in place if they are likely to occur in the tribosystem of interest. If they are not part of the system of interest, they should be removed by cleaning. Valid wear results require that a cleaning procedure be used that will remove contaminants and not leave another contaminating film. The cleanest surface is the as-machined or as-ground surface (without coolants and the like). These types of surfaces prevent cleaning-solution contamination problems. If metal surfaces must be cleaned of contaminating films, then a refluxed solvent system should be used. Plastics are difficult to clean with solvents because of the risk of chemical alteration of the surface. Commercial glass-cleaning solutions have been employed to clean plastic test surfaces with favorable results. Ceramics can also be cleaned by these solutions. Sometimes even water-based cleaners can chemically alter a ceramic surface. The safest technique for removing contaminants on ceramics is dry lapping or similar abrasive finishing.

Absorbed films from the atmosphere can be another source of contamination. To cite an example, many plastics were tested in a comprehensive laboratory as a track material for conveyance of plastic parts. Test results indicated that an ultrahigh-molecular-weight polyethylene had the lowest friction. Hundreds of feet of this track were installed in the manufacturing plant, but within a month the track was pulled out and replaced because of an environmental factor that had not been considered during laboratory testing. Although there were no problems with the track during the week when there was a three-shift operation, after the weekend, parts would stick to the track with great frequency. Only after several hours downtime and frequent "nudging" of the parts would the system begin working properly. Further study indicated that when the track material was allowed to sit idle for a period of 48 h, some type of film formed that deleteriously altered the friction characteristics.

This example points out the importance of carefully duplicating the system environment, even to the extent of allowing surfaces to sit idle in production environments. Many manufacturing plants have an atmosphere that can leave a condensate on a surface. If this is likely to occur in a tribosystem under study, then it should be made part of the friction testing.

Test Parameters

Friction test parameters, such as speed and load, should be as close as possible to actual conditions. Sometimes, however, it is difficult to simulate these parameters in a laboratory environment. The speeds and loads may be out of the range of the lab equipment, or testing at actual conditions may take too long. A valid friction test does not require that speeds and loads be exactly duplicated. Friction coefficients are relatively insensitive to speed and loads up to the point where these parameters affect the properties of the test surfaces. Speed will affect friction results on most metals when it is sufficiently high to soften the surface and alter mechanical properties. With plastics, this temperature may be only a few hundred degrees Fahrenheit. With metals, the critical temperature is close to the stress-relief temperature for a given alloy.

Sliding speed for a friction test can range from 10 mm/s (0.4 in./s) for an inclined plane test to 100 m/s (325 ft/s) for high-speed yarn friction tests. For optimum results, test equipment should yield a wide range of sliding speeds. If the actual operating speed is unusually high or low, unique friction conditions may be produced that can be simulated only by test speeds in the range of concern. For example, a laboratory study was conducted on plastic and competitive plain bearings for a compact-disk drive scanner that had to slide (bushings versus hardened steel shafts) in a linear motion at a speed of 1.6 μm/s (0.00006 in./s). This slow speed caused many materials to display stick-slip behavior. When higher speeds were used, this did not occur. Under such conditions, the speed of the tribosystem should be duplicated as closely as possible.

The same type of reasoning should be applied to test loads. If a system under study uses loads of only a few grams, this should be duplicated. If the system under study is subject to extreme forces, such as those that occur in threaded fasteners, there is little recourse other than to use these very high forces. When studying bolt friction, most investigators use actual bolts and nuts and a washer equipped with a strain gage for measurement of bolt tension. This is because of the difficulty in obtaining a bench-top friction tester that can apply a 45 kN (10,000 lbf) normal force.

As a rule of thumb, if it is not possible to simulate the service load, then the highest load that does not exceed the compressive strength of either member in the test couple should be used. This is particularly applicable to Hertzian loading situations. If the compressive strength of the material is exceeded, the friction test only yields the forces to plastically deform the surfaces. Test speeds and loads that produce significant wear should be avoided unless that is part of the tribosystem under study.

As previously mentioned, friction force may not be constant throughout a test, and questions may arise as to the value that should be reported. The most meaningful data are the minimum, the maximum, the mean, and the standard deviations. This allows the use of statistical analysis to determine wether friction coefficients for various systems are really different. Erratic systems such as that described in Fig. 12(c) will be treated with statistics. In friction tests were the goal is to find a couple with the smoothest motion, a suitable test parameter for ranking materials is the standard deviation of the friction force. This shows the variability of the friction forces for the various systems. In fact, for all friction tests it is recommended that statistical tests be used. The simplest test of differences is illustrated in Fig. 14. If the mean of test data is plotted with error bars coinciding with plus or minus two standard deviations, a visual test of differences can be made. If the error bars overlap, the test results are not statistically different. In Fig. 14, couples B, C, and D have the same frictional characteristics; only couple A differs.

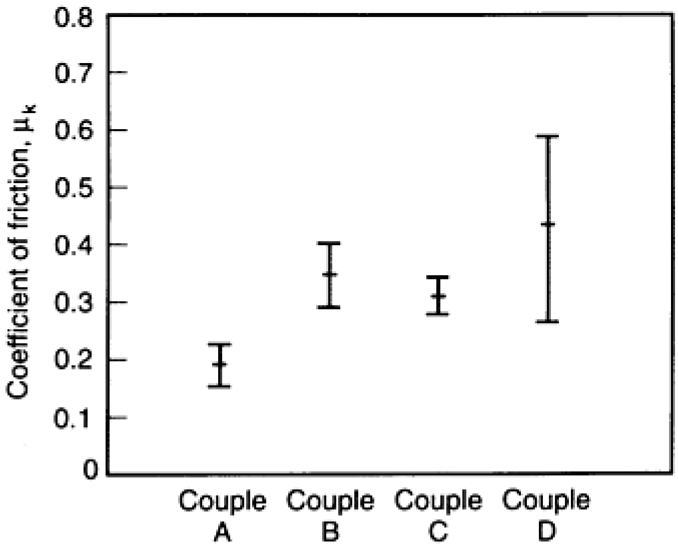


Fig. 14 Graphical use of statistics to show differences in friction coefficients between various material couples

Sample size must also be considered in statistics tests. There are a number of ways to calculate sample size, but most statisticians prefer a large number (for example, 40) of replicate tests. Because most friction tests are very repeatable within a laboratory, this number may be a bit high. Most studies using a bench-top friction tester can have statistical accuracy with as few as 10 replicates per system.

Friction Measurements

There are a number of precautions that must be observed during friction measurements, and most have to do with the measuring equipment. Some of the more commonly used force-measuring systems are illustrated in Fig. 15. Probably the most important consideration with regard to instruments for measuring friction forces is selection of a force-measuring system that is suitably sized for the task. If the tribosystem being tested will produce friction forces of about 4.5 N (1 lbf), it is recommended that a force-measuring system with a capacity of only about 45 N (10 lbf) be used. In other words, a sensor with a capacity not more than 10 times the force to be measured should generally be selected for a given application.

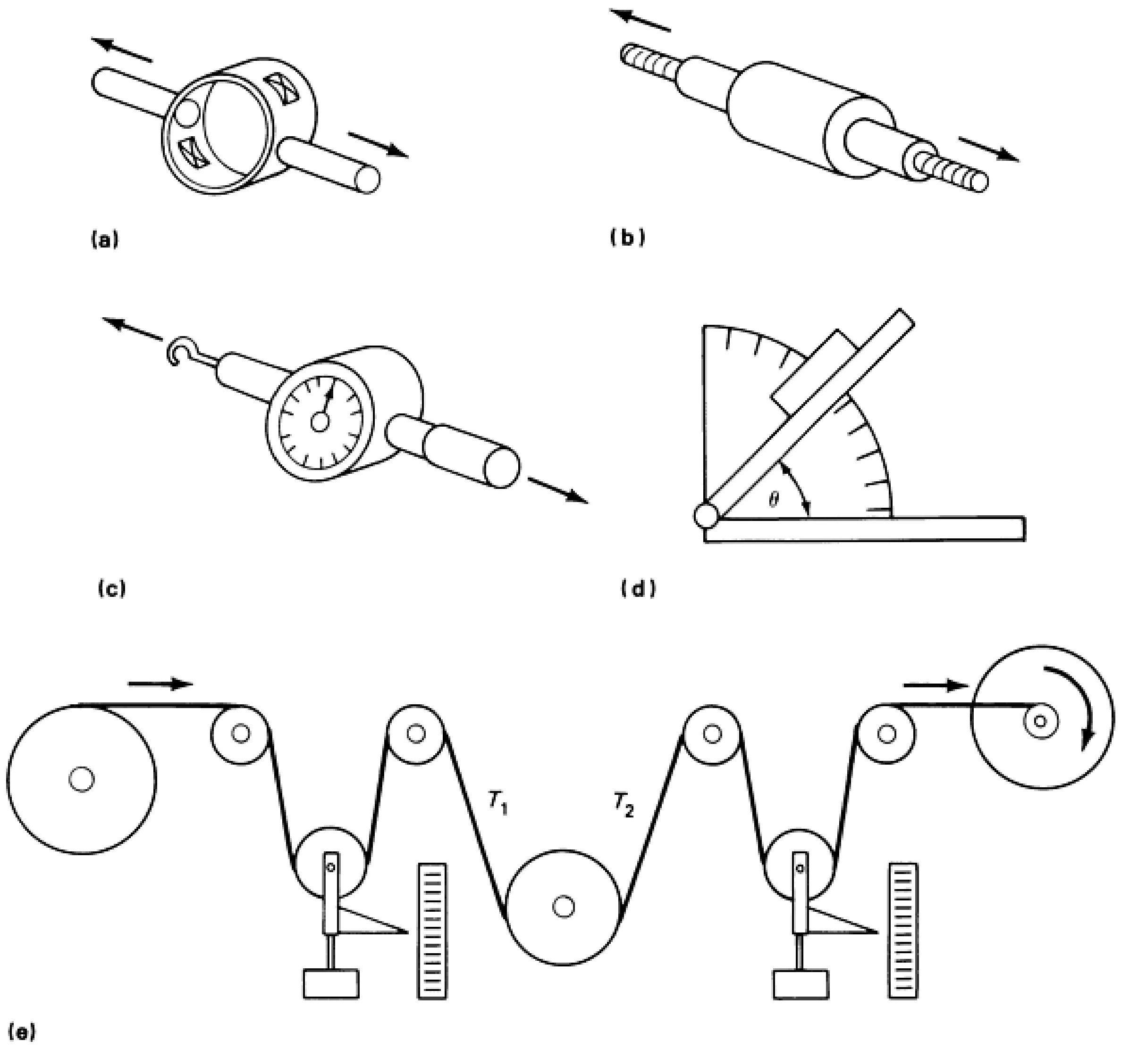


Fig. 15 Friction force sensors. (a) Strain ring with strain gages. (b) Strain gage load cell. (c) Mechanical force gage. (d) Inclined plane. (e) Web tension tester (tension is measured by the position of the weighted sheaves)

The strain ring illustrated in Fig. 15(a) is assembled in the laboratory and features four strain gages adhesively bonded to the ring to form a Wheatstone bridge strain rosette. This type of system can be very accurate, because the ring can be made a heavy or as light as desired. The disadvantages are that such gages are more fragile than purchased gages, and humidity and temperature can affect the adhesion of the gages. This system also requires assembly by a person with some expertise in strain gages.

The commercially available strain-gage load cells shown in Fig. 15(b) are quite durable. Most have some type of internal diaphragm that is instrumented with stain gages. However, these devices can be destroyed by a momentary load greater

than its capacity, necessitating expensive replacement. Many companies offer these gages with mechanical stops to prevent overload damage.

Mechanical force gages (Fig. 15c) work much like strain gages, except that force measurements are made by a spring and sliding system arrangement. The greatest disadvantage of such devices is that they have no recording capability. They are also prone to inaccurate readings by operators.

Friction coefficients measured on inclined planes (Fig. 15d) require only the measurement of the angle at which motion of the test block occurs. The most common technique for measuring this angle is the simple incorporation of a protractor scale in the device. A more accurate system uses an electronic level on the inclined plane. These devices have digital readout capability and are much more accurate than visual assessment of the angle of an inclined plane. One precaution in using this type of device is that the speed at which the plane is raised will affect the breakaway angle. The more reliable inclined plane rigs have a motorized raising device on the table.

The device illustrated in Fig. 15(e) is commonly used to measure the friction coefficient of a moving web over a stationary surface or over a roll where relative slip is occurring. The tension in the web on either side of the test roll will yield the coefficient of kinetic friction when the capstan formula is used. The tension-measuring roll, called a "dancer roll," moves up and down with changes in web tension. The disadvantage of these systems is that they are not very easy to gage electronically, and thus continuous recording mae be difficult. Most dancer rolls are being replaced by cantilever rolls with strain gages incorporated in their mounting brackets.

Whichever force-measuring technique is used, it is most important that the device be sized correctly and that friction forces be recorded over a sufficient period of time to ensure that steady-state conditions exist.

If an electronic device is used to measure and record friction force, the recording system is of concern. The current trend is to use computers to record friction and wear results from test machines. When friction force is digitized, however, the electronic force sensors do not miss any forces; the resulting "hash" in the friction recording may make it difficult to measure averages and ranges and other parameters for statistical data analysis. Although most systems can be electronically filtered, the consensus at the present time is that digitizing of friction forces in friction or wear tests creates too many complications. The filters in strip-chart recorders apparently provide adequate noise filtering and thus reasonable measurements of friction force.

The final precaution that warrants consideration is the system stiffness. If a system is prone to stick-slip behavior or high static friction, a highly compliant force-measuring system should be used. From a practical standpoint, an elastic force-measuring system usually consists of nylon fishing line to pull member A along member B. If a stiff-movement system such as a screw drive is used to produce specimen motion, phenomena such as stick-slip behavior may not be produced. The decision as to whether a high-stiffness (steel) or low-stiffness (plastic) system should be used must be based on the intended application. If the tribosystem of interest is highly elastic, the plastic-movement system should be used. If the system is rigid, the screw-drive system (or a similar system) should be used.

Reporting System Losses

A proposed method for analyzing wear problems is to look at the tribosystem and tabulate all of the system inputs, outputs, and disturbances (Fig. 16). Friction is considered to be a system loss, but certain components of friction should be noted and recorded. From the aspect of system usability, sometimes these system losses are more important than the friction force. For example, in a study on the wear of various cemented carbide couples, it was learned that some carbide couples produced squealing. The friction coefficient was slightly higher than comparable metal-to-metal couples (about 0.5 compared with 0.4) and the wear was very low, but this couple was deemed unacceptable for use because of the noise.

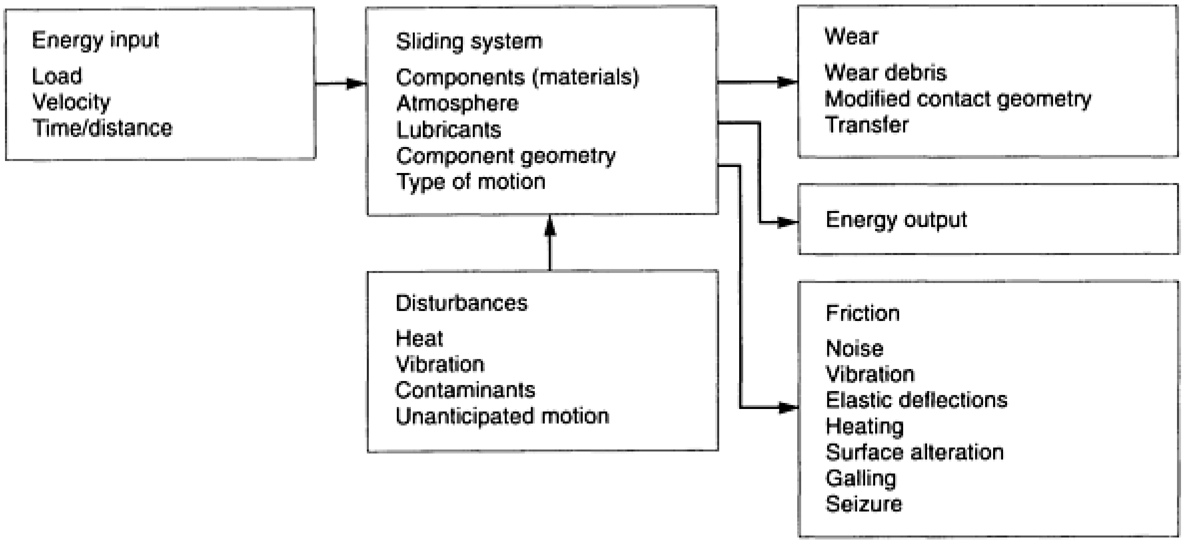


Fig. 16 Systems approach for analyzing friction and wear problems

In Fig. 16, important losses include vibration, elastic deflections, heating, surface alteration, galling, and even seizure. Vibration as an output of a sliding system often indicates that stick-slip behavior is prevalent. Noise is often the result of stick-slip behavior, but a system that exhibits such behavior does not necessarily emanate noise. Although vibration may not be apparent on friction force traces, it may show up on deflection or displacement transducers. This type of behavior from a sliding system is usually undesirable. Accurate measurement of vibration may require the use of accelerometers on one of the members of the sliding system.

Likewise, elastic deflection, which can occur at a sliding interface, may not show up on the friction force recordings; this deflection may mean that the couple under study has unacceptable frictional characteristics. For example, when several elastomers that were undergoing friction tests were slid on a paper counterface, they bent over in the direction of motion; the contact geometry was changed from the point contact of a hemispherical rider to a line contact of a bent hemispherical-ended rod. The friction force was apparently high enough to cause this deflection.

Although heating is an obvious result of friction between sliding members, it is often not measured. The temperature rise is often significant, and it is easy to measure. The mechanical properties of plastics are susceptible to degradation by heating to relatively low temperatures. The temperature rise at a sliding interface is the result of the properties of the materials in contact in addition to the sliding conditions. It will be different for different couples that may have the same friction coefficient. Therefore, for sliding systems that may be affected by frictional heating of the interfaces, a valid friction test should record the temperature rise.

Surface alteration is another important aspect of many wear and friction tests. Whenever wear occurs in a sliding test, the friction coefficient is not that of the test couple alone, but it is the system that comprises the couple as well as wear debris in the interface. When wear testing couples that are not supposed to wear during friction testing, it is important to examine both surfaces for alterations. Damage often may be caused by polishing or scratching. When friction alters the prevailing surface texture, a wear test has been performed, not a friction test. The friction force measured and the coefficient of friction must be reported for a worn surface.

Galling and seizing are the worst possible results of a friction test. Galling is characterized by the formation of microscopic cumulative material transfer during sliding, and seizure (stopping of motion) can be the net result. If a couple seizes, there will be no friction coefficient to report, but merely the fact that the couple seized. If galling occurs, the friction force will often decrease (Fig. 17), but the surfaces will be damaged. This can produce data that misleads a user, who may think that the couple works fairly well because the friction coefficient was low, when actually galling occurred and the material couple is not compatible.

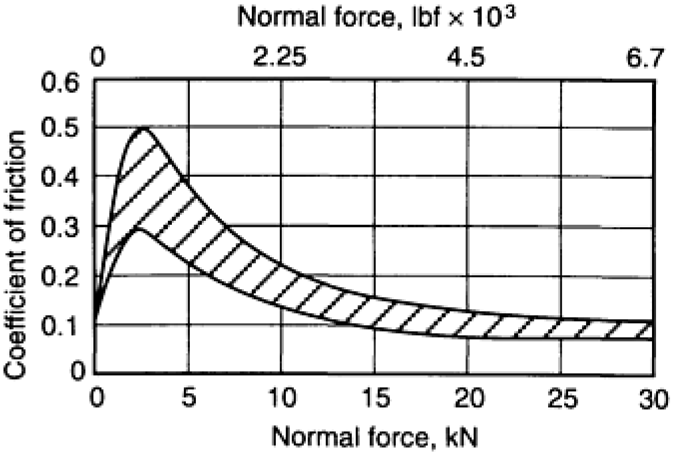


Fig. 17 Typical friction coefficients derived from galling tests (various metal/metal couples)

Friction Databases

The first friction database was compiled by J.T. Desagulies around 1725 (as reported by Dowson in Ref 2). Desagulies tabulated the friction coefficients for the couples of interest at the time. Current handbooks tabulate data for 50 or so materials, with limited documentation regarding test conditions. These tabulations are of little use if the application requires knowing a friction coefficient within an accuracy of ±0.2. Differences in the tribosystem used to make the measurements can, for example, produce a result of 0.1 for a couple on device No. 1 and a result of 0.3 for the same couple tested on device No. 2. It can be stated with a high degree of confidence that measuring techniques will have a significant effect on the friction coefficient of a particular couple in the unlubricated condition. Differences may exist in lubricated systems, but the coefficients will be much lower and an accuracy of ±20% results in a much smaller number. For example, well-lubricated steel couples may have a kinetic coefficient of friction of 0.05 (±20% makes the number 0.04 to 0.06). In other words, existing friction databases have limited utility unless the test conditions used to develop the data are stated and the application conditions are similar.

In order to determine how friction databases should be formulated and used, ASTM Committee G-2 on wear and erosion began developing a standard format for friction databases in 1987. Although this work is ongoing, progress has been made as to the type of data that should go into databases. The minimum results to be reported are:

- Test couple (member 1 and member 2)
- Static coefficient of friction
- Kinetic coefficient of friction

The minimum test condition information includes:

- Apparent pressure
- Normal force
- Velocity
- Temperatures (bulk) of samples
- Test atmosphere
- Lubricant
- Sliding distance (when μ was measured)

Other types of data would also be desirable, but even the ASTM-recommended list is difficult to deal with in database or spreadsheet types of software. The strategy is to have these data in a database so that selective data can be tabulated (see

Table 3). The motivation for establishing friction databases is the elimination of repetitive tests. Even within a single laboratory, it is not uncommon to see the same couples brought in for study several times over a period of several years. Without a database, the tests are rerun each time. The long-range goal is to have published data that can be used by design engineers in the same way that designers use corrosion data generated by countless sources over many years.

Table 3 Friction and wear data of selected plastics tested against polycarbonate containing 12% polytetrafluoroethylene

Plastic[(a)]	Kinetic coefficient of friction, μ_k	Specific wear rate of selected plastic, $K \times 10^{-6}$[(b)]
PC	0.19	7.00
PC + PET	0.19	7.26
PCTG + 30% GF	0.36	9.00
PET + 30% GF	0.27	9.00
PET + 30% GF + mica	0.29	8.00
PC + 10% aramid	0.09	4.00
PA + 10% aramid	0.08	0.80
PA + 15% aramid + 10% TFE	0.07	0.50
LCP wear grade	0.10	2.00
LCP + mineral	0.07	0.60
PC + 40% aramid	0.18	9.00
PC/IPN + 2% aramid	0.15	2.00
PC + 20% aramid	0.20	7.00

(a) PC, polycarbonate; PTFE, polytetrafluoroethylene; PET, polyethylene terephthalene; GF, glass fiber; PA, polyamide; LCP, liquid crystal polymer; IPN, interpenetrating networks; TFE, tetrafluoroethylene.

(b) Parameters: friction force, *F,* 9.86 N (2.20 lbf); velocity, *V,* 0.208 m/s (0.682 ft/s); sliding distance, *D,* 732.0 m (2402 ft)

References

1. A.Z. Szeri, *Tribology: Friction, Lubrication and Wear,* Hemisphere Publishing, 1980, p 2
2. D. Dowson, *History of Tribology,* Oxford University, 1979, p 22-23
3. K.C. Ludema, Friction, A Study in the Prevention of Seizure, *ASTM Stand. News,* May 1987, p 54-58
4. J.J. Bikerman, *Polymer Friction,* Plenum Press, 1974, p 149
5. F.B. Bowden and D. Tabor, *The Friction and Lubrication of Solids,* Clarendon Press, 1950
6. M.J. Neale, Ed., *Tribology Handbook,* John Wiley & Sons, 1973, p C-7
7. J.A. Schey, *Tribology in Metalworking,* American Society for Metals, 1983
8. D. Whitehouse, Friction and Surface Measurement, *Surf. Topog.,* Vol 1, 1988, p 427-433
9. I.V. Kragelskii, The Nature of Polymer Friction, *Polym. Mech.,* Vol 8 (No. 5), Sept 1972, p 699-707
10. D. Tabor, *Proc. R. Soc. (London) A,* Vol 255, 1989, p 378
11. I.V. Kragelskii, *Friction and Wear,* Butterworths, 1965
12. K. Friedrich, *Friction and Wear of Polymer Components,* VDI-Verlag, 1989, p 14
13. D.H. Buckley, *Surface Effects in Adhesion, Friction, Wear and Lubrication,* Elsevier, Amsterdam, 1981, p 915
14. "Standard Terminology Relating to Wear and Erosion," G 40, *Annual Book of ASTM Standards,* ASTM
15. K.G. Budinski, Friction of Plastic Films, *Wear of Materials 1989,* American Society of Mechanical Engineers, 1989, p 459-468
16. Y.-R. Jeng, Experimental Study on the Effects of Surface Roughness on Friction, *Tribol. Trans.,* Vol 33 (No. 3), 1980, p 402-410

Friction during Metal Forming

Betzalel Avitzur, Metalforming Inc.

Introduction

FRICTION exists in any metal-forming process. Whenever two solid surfaces are in contact and relative motion, a resistance (friction) to this motion arises. Friction is the last frontier in the study of metal forming. For example, in the process of wire drawing, independent parameters such as reduction and die angle can be measured directly. Friction, however, is not directly measurable, nor is it really an independent parameter. Yet in many metal-forming processes, the effect of friction is as important as that of measurable independent parameters.

During wire drawing, a wire slides over the conical and cylindrical surfaces of a die (Fig. 1). If no lubricant is used, there is direct metal-to-metal contact between the wire and the die. The pressures between the die and the wire are very high (approximately equal to the flow strength, σ_0 of the wire). The relative motion, together with high pressure and high friction resistance, results in the generation of heat. The relative movement of the mating surfaces causes them to be damage by wear. Buildup of foreign matter over the surface of the die is also possible.

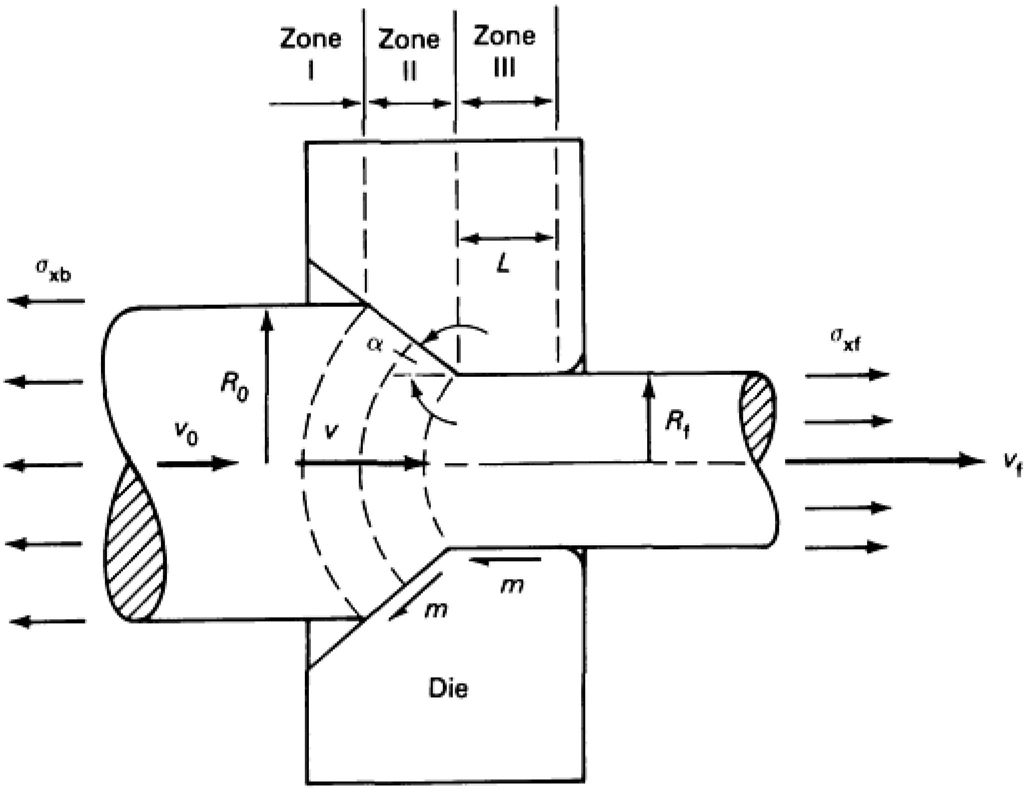


Fig. 1 Schematic of wire drawing or extrusion processes

No surface is geometrically perfect. Surfaces contain irregularities that form peaks and valleys. Thus, contact between the die and the workpiece is maintained over limited portions of the apparent interface. The apparent area of contact is the total area, but the actual area of contact is limited to that between the peaks of the opposing asperities (Fig. 2). If the pressure (*p*) is defined as the total force (*W*) divided by the apparent cross-sectional area (*A*), the local pressure at the points of contact can be much higher (Ref 1). The asperities flatten under the pressure, and the area of flattened asperities adjusts itself to carry the load by plastic deformation of the asperities, even when the bulk of the workpiece is in the elastic state.

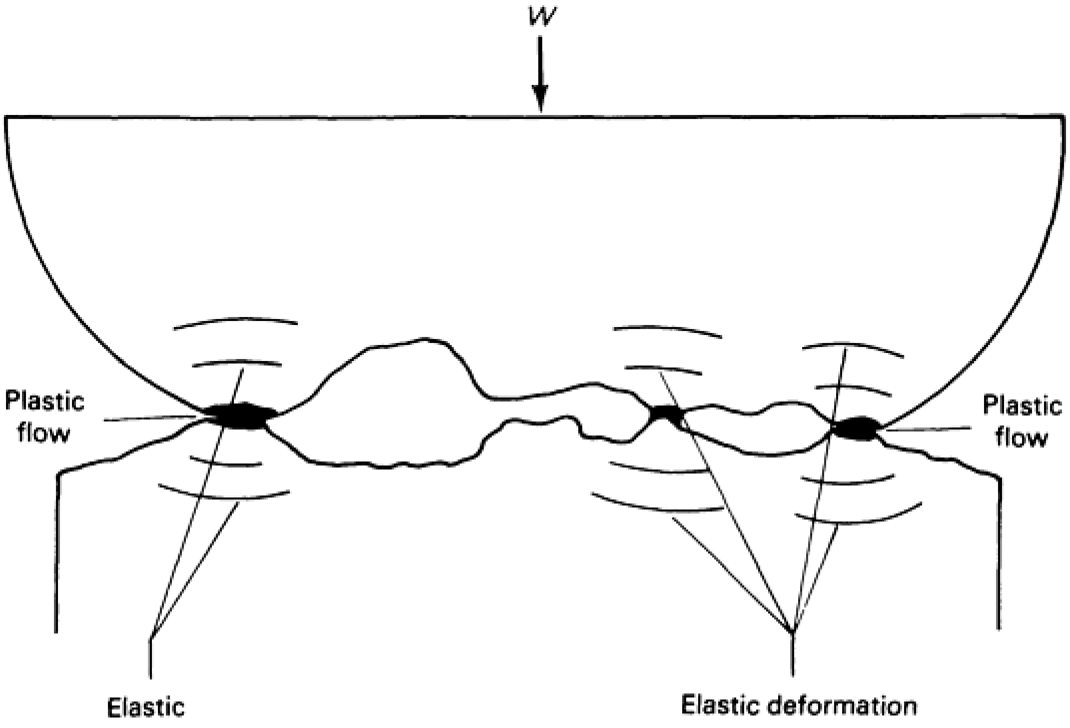


Fig. 2 Schematic of surface irregularities

Surface irregularities and their behavior during sliding, together with the lubricant and surface chemistry, are key factors in the characterization of friction and wear. Figure 3 illustrates some of the various modes of asperity deformation behavior. The steady-state wave model, which is shown schematically in Fig. 3(b), is described in the Appendix to this article. This model provides explicit expressions for the characterization of friction resistance as a function of pressure (Fig. 4a), speed and Sommerfeld number (Fig. 4b), and surface geometry. Terms commonly used in metal-forming operations are defined in Table 1.

Table 1 Nomenclature for friction in metal-forming processes

A	**Cross-sectional area**
f%	Percent forward slip in strip rolling
ℓ_0	Length of the asperity
L	Length of the bearing (land) of the die
m	Constant shear, or friction factor
p	Pressure
r%	Percent reduction in area
R_0, R_f	Original and final radius of a wire
R_O	Radius of the roll in strip rolling
R_o, R_i, R_n	Outer, inner, and neutral radius, respectively, of a deforming ring
S	Sommerfeld number
t_0, t_f	Original and final thickness of the strip
T_0	Thickness of a deforming ring
v_0	Sliding velocity
W	Total normal load
$\dot{W}_f$	Friction power losses
$\dot{W}_i$	Internal power of deformation
$\dot{W}_s$	Shear or redundant power of deformation
α	Semicone angle of the die
α_1	Angle of inclination of the asperity (the wedge)
α, α_n, α_2	Angular positions, respectively, of: an arbitrary point; the neutral point; and the point of contact

$\dot{\gamma}$	Shear strain rate
ϵ	Thickness of the film of lubricant
η	Viscosity
μ	Coulomb's or Amonton's coefficient of friction
$\sigma_{\mathbf{0}}$	Flow strength
$\sigma_{\mathbf{s}}$	Shear stress in the liquid
$\sigma_{\mathbf{xb}}$	Extrusion pressure with its negative sign (that is, $\sigma_{xb} = -p$); also back tension
$\sigma_{\mathbf{xf}}$	Drawing stress or front tension
τ	Shear stress
$\Delta \boldsymbol{T}_{\mathbf{0}}$	Reduction in thickness

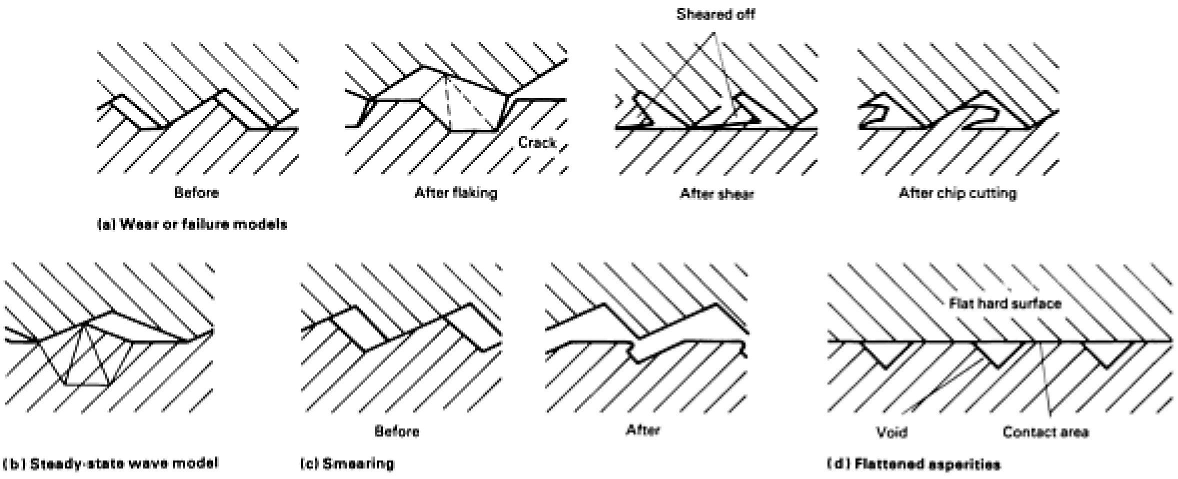


Fig. 3 Schematic of asperity deformation behavior

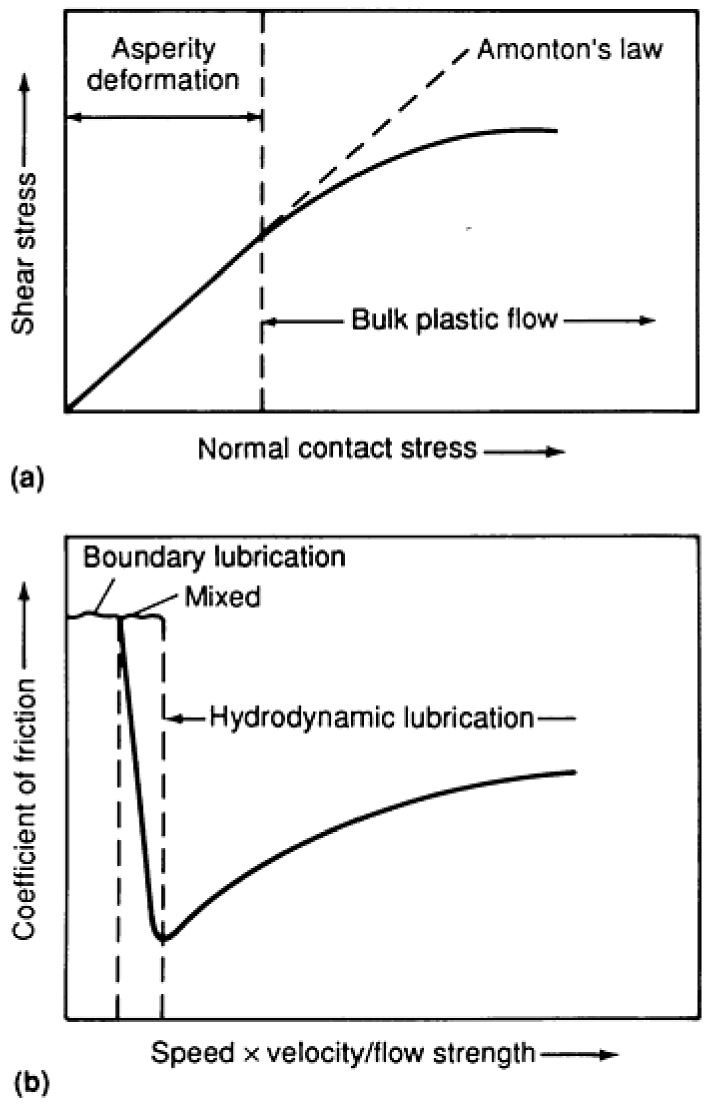


Fig. 4 Characterization of friction resistance as a function of (a) pressure, *p*, and (b) Sommerfeld number, *S*

Modeling of Friction

The difficulties in the determination of the friction value lie in the complexity of the phenomena (Fig. 4 and 5, and Appendix) and in the inability to accurately measure shear stresses. Therefore far-reaching approximations, as will be described presently, are used to describe friction behavior during metal forming. These approximations deal with apparent friction rather than with the fundamental phenomenon. One of the consequences of this approach is that friction must be measured separately for each forming process. Under presumably identical conditions of surface finish and lubrication, wire drawing will produce different friction values than strip rolling. As a result, when friction during rolling is to be determined, the rolling process must be simulated. The same holds true for the other metal-forming processes.

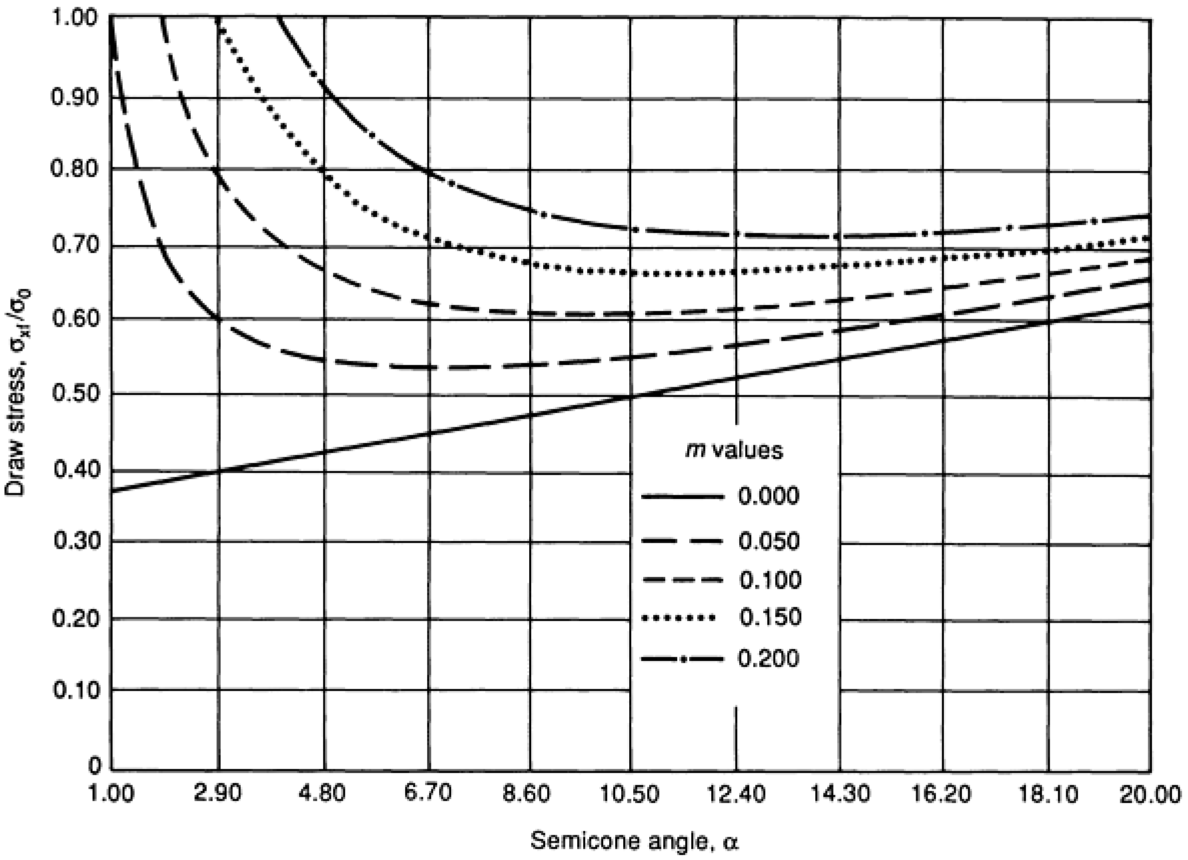


Fig. 5 Relative drawing stress as a function of semicone angle and friction factor (*m*). See also Eq 4.

Standard practice in the study of metal forming is to assume that the resistance to sliding along the interface between the tool and the workpiece is uniform along the entire contact surface. The most common simplifying assumptions made with regard to friction stress (τ) between the workpiece and the tool involve Coulomb friction, constant friction, and hydrodynamic-, hydrostatic-, and thick-film lubrication.

For Coulomb friction, it is assumed that the shear stress (τ) is proportional to the pressure (p) between workpiece and the die. It follows then that:

$$\tau = \mu p \qquad \textbf{(Eq 1)}$$

where the proportionality factor (μ) is called the Coulomb coefficient of friction.

For constant friction, it is assumed that the shear stress is proportional to the flow strength of the workpiece material, and

$$\tau = m\sigma_0/\sqrt{3} \qquad \textbf{(Eq 2)}$$

where the proportionality factor (m) is called the shear (or friction) factor, with $0 \leq m \leq 1$. The factors (μ) and (m) are assumed constant for a given die, workpiece, and lubricant.

When a lubricant film separates the workpiece from contact with the die, hydrodynamic or hydrostatic film lubrication prevails together with its special laws of shear within the lubricating medium. Sometimes high-viscosity lubricants adhere to the workpiece, resulting in thick-film separation of the workpiece from the tool. Film lubrication may also separate the workpiece from the die on the entry side to a smaller or larger extent. At the extreme, the entire workpiece is separated

from the die by this lubricant film. Under such conditions, the parameters μ or *m* are replaced by the viscosity of the lubricant (η), where stress in the liquid (σ_s) is expressed as

$$\tau = \sigma_s = \eta\dot{\gamma} \qquad \textbf{(Eq 3)}$$

where $\dot{\gamma}$ is the shear strain rate within the lubricant.

In the section of this article on "Measurement of Friction," the determination of friction is described for forging, wire drawing, and strip rolling. For each of these processes, the apparent friction is determined experimentally through the application of analytical solutions. The experimental data is treated by the mathematical expression of the relation between the parameters that were measured and the sought friction. In each technique a minimal use of instrumentation is required. All these assumptions for the characteristics of friction--namely Coulomb's/Amonton's coefficient (μ), constant factor (*m*), and film lubrication (η)--are treated. An iterative procedure can also be implemented when friction and pressure depend on each other, are solved simultaneously, and their distributions along the contact surface are treated as variables. Bay (Ref 2) gives such a treatment for the extrusion process.

Modeling of Flow through Conical Converging Dies

In the process of wire drawing, a wire is pulled through a converging die where its size is reduced from R_0 to R_f (Fig. 1). Passing through the die, the wire rubs against the conical and cylindrical surfaces of the die and encounters friction resistance. The effect of friction on the drawing force and drawing stress during the fabrication of wire through conical converging dies is discussed in Ref 3, 4, 5, 6, 7, 8, 9, 10, 11. The characteristics of the die and the flow patterns in Fig. 1 are common to wire drawing, open-die extrusion, and hydrostatic extrusion. A typical solution presented in Ref 12 expresses the relative drawing stress (σ_{xf}/σ_0) as a function of the input parameters, including the constant friction factor (*m*), as follows:

$$\frac{\sigma_{xf}}{\sigma_0} = \frac{\sigma_{xb}}{\sigma_0} + \dot{W}_i + \dot{W}_s + \dot{W}_f \qquad \textbf{(Eq 4)}$$

where σ_{xf} and σ_{xb} are the front and back tensions, respectively, and where

$$\dot{W}_i = 2\,f(\alpha)\,\ln\frac{R_0}{R_f}$$

is the contribution of the internal deformations,

$$\dot{W}_s = \frac{2}{\sqrt{3}}\left(\frac{\alpha}{\sin^2\alpha} - \cot\alpha\right)$$

is the shear or redundant power term,

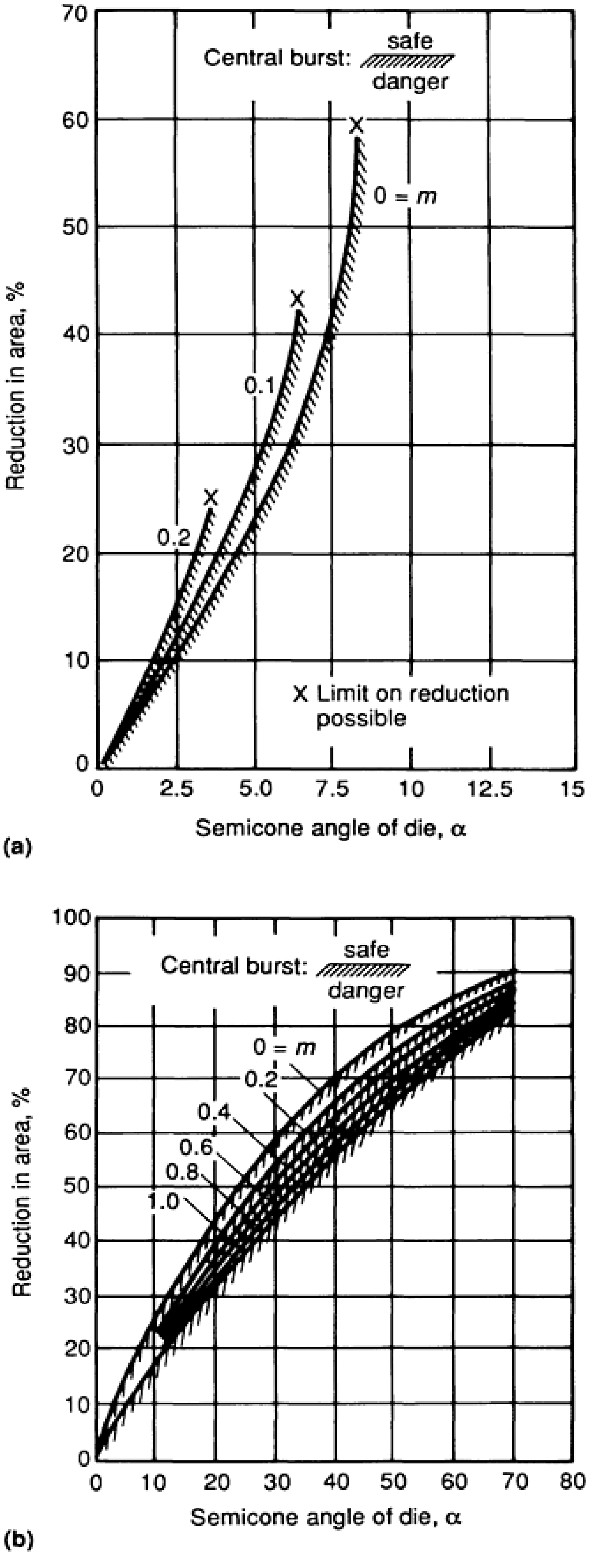
$$\dot{W}_f = \frac{2}{\sqrt{3}}\,m(\cot\alpha)\ln\frac{R_0}{R_f}$$

is the friction term along the conical surface of the die, α is the semicone angle of the die, and

$$f(\alpha) = \frac{1}{\sin^2 \alpha} \left\{ 1 - \cos \alpha \sqrt{\left(1 - \frac{11}{12} \sin^2 \alpha\right)} + \frac{1}{\sqrt{(11 \cdot 12)}} \ln \frac{1 + \sqrt{11/12}}{\sqrt{11/12} \cos \alpha + \sqrt{(1 - (11/12) \sin^2 \alpha)}} \right\}$$

The characteristics of Eq 4 are presented in Fig. 5, where the semicone angle of the die (α) is the abscissa and the relative drawing stress the ordinate. Each curve in Fig. 5 demonstrates a minimum at some optimal angle. For angles smaller than the optimal angle, friction losses ($\dot{W}_f$) are excessive. Friction losses drop with increasing die angles, but redundant power losses increase with increasing die angle. Beyond the optimal angle, excessive distortion occurs with increasing die angles, and the drawing stress increases. The optimal die angle that minimizes the drawing stress is increasing with the increase in reduction and friction. With higher friction it is advisable to use larger die angles.

More expressions have been derived for the determination of process limitations due to tearing, dead zone formation, and shaving (Ref 12). When the phenomenon of central burst is analyzed, one finds that increasing friction deters central burst during extrusion, but promotes it during drawing (Fig. 6).



(a)

(b)

Fig. 6 Criteria for central burst in (a) drawing and (b) extrusion

Hydrodynamic Lubrication. The analysis of hydrodynamic lubrication by Hillier (Ref 4) offers a simple and applicable treatment. This treatment, with a slight modification to account for variations in the thickness of the lubrication film, is presented in Ref 12. The historical development and the state of the art in the application of hydrodynamic lubrication to wire drawing are described in Ref 5 and 6.

In early studies of the equipment design for analysis of hydrodynamic lubrication (Ref 7, 8), a long tube with a narrow gap between the tube and the wire was firmly attached at the entrance side of the die (Fig. 7). The lubrication adhered to the wire and was dragged into the clearance between the tube and wire. At about 3 m/s (600 ft/min), the pressure that built up at the approach to the die reached 70 to 275 MPa (10 to 40 ksi), and the liquid formed a film between the wire and the die (Ref 7).

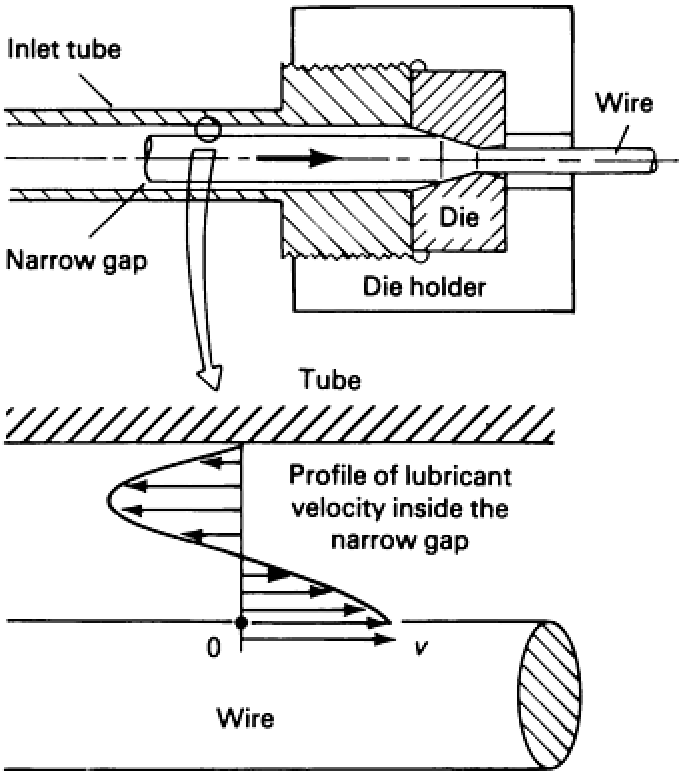


Fig. 7 Schematic of inlet tube for hydrodynamic lubrication

A method whereby the wire passes through a pressurized chamber (Fig. 8) has been studied by several investigators (Ref 9, 10, 11). Using this technique, the lubricating liquid is supplied to the chamber at high pressure by external means (Ref 13). The hydrodynamic effect does not depend entirely on the speed, viscosity, or chamber/tube length. With higher pressure in the chamber, hydrodynamic lubrication can be accomplished with lower velocities and lower fluid viscosities.

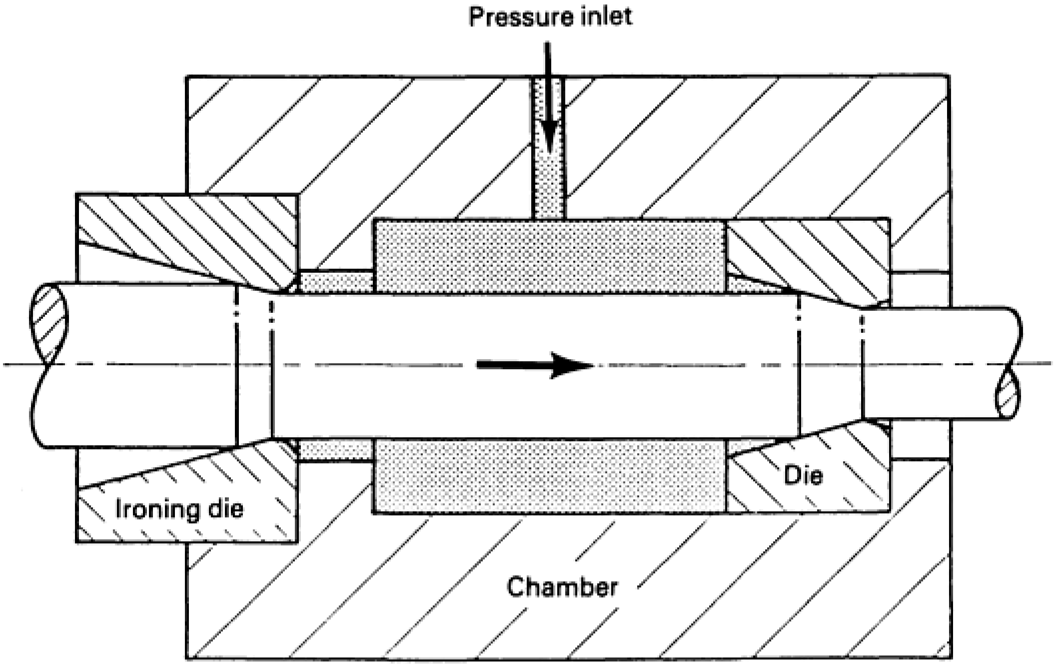


Fig. 8 Schematic of wire drawing through a pressurized chamber

The introduction of hydrodynamic lubrication through a pressure box is shown in Fig. 9. The pressure chamber is inserted into the soap bin of a conventional wire-drawing bull block. The bin preceding the entrance die to the chamber is filled with powdered soap. The wire running through the powder supply drags powder into the chamber through the narrow gap between the incoming wire and the approach die.

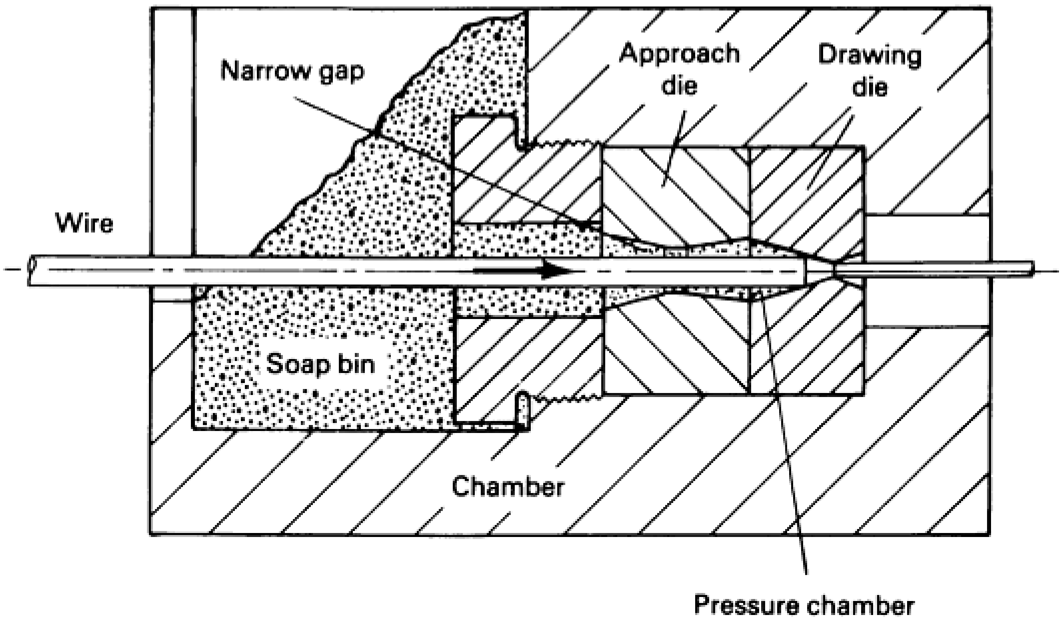


Fig. 9 Set-up for hydrodynamic lubrication with dry soap

When the wire is pulled through the soap, some soap is dragged into the chamber. In the small clearance gap, the soap powder shears, heats because of this shear, and then melts. Liquid lubricating soap then enters the pressure chamber. The faster the wire is drawn, the higher are the temperatures and pressures introduced in the chamber causing hydrodynamic

lubrication. A chemical bond between the wire and the molten soap is also produced. As soon as the wire exits the die, its temperature drops and the soap freezes and forms a layer on the wire. There is now a layer of lubricant performing the duties of hydrodynamic lubrication, that is, keeping full separation between the deforming wire and the next five or six dies. The pressures produced in these chambers, without any external or auxiliary agent, are in excess of 275 MPa (40 ksi).

One difficulty associated with the use of a powdered-soap bin is in maintaining steady drag of powder into the chamber. Inconsistencies in the particle size and dryness of the powder vary the effectiveness of its adhesion to the wire and the quantity of powder dragged. At the extreme, a hollow channel in the powder surrounds the incoming wire and no soap is dragged in. Greater uniformity of the powder and constant agitation (directly or through the box) may improve the performance. Proper surface preparation also may be called for. A recent development utilizes the conventional powder spray system with electrostatically charged particles.

With phosphating, the most popular method, a predetermined layer of a spongy phosphate coating is applied to the surface of the wire electrolytically. This sponge can absorb and retain a large volume of lubrication liquid. Even without the lubricant, the phosphate sponge provides an effective separation between the die and the workpiece, thereby minimizing friction and wear.

The average shear stress and the thickness of the lubricant (Eq 9.19 of Ref 12) are:

$$\frac{\bar{\tau}}{\sigma_0} = 2\,\frac{(\cos^2\alpha/\sin\alpha)\,(1 - R_f/R_0) + L/R_f}{(1/\sin\alpha)\,[(R_0/R_f)^2 - 1] + L/R_f} \times \frac{\left[\dfrac{2v_f}{R_f\sigma_0}\right] f(\alpha)}{[(\cot\alpha\,\cos^2\alpha)/3]\left[1 - \left(\dfrac{R_f}{B_0}\right)^3\right] + \dfrac{L}{R_f}} \qquad \textbf{(Eq 5a)}$$

and

$$\frac{\epsilon}{R_f} = \sqrt{\left(\frac{\eta v_f}{R_f\sigma_0}\right)\frac{1}{f(\alpha)}\left\{\frac{\dfrac{\cot\alpha\,\cos^2\alpha}{3}}{\left[1 - \left(\dfrac{R_f}{R_0}\right)^3\right] + \dfrac{L}{R_f}}\right\}} \qquad \textbf{(Eq 5b)}$$

When the dimensionless factor $\eta v_f/(R_f\sigma_0)$ (called the Sommerfeld number) increases from zero, both the relative average shear stress (τ/σ_0) and the relative thickness of the lubricant (ϵ/R_f) increase. For small values of the Sommerfeld number, the expected thickness of the lubricant is lower than the surface roughness of the material and the die, and hydrodynamic lubrication does not prevail. Therefore, for small values of $\eta v_f/(R_f\sigma_0)$, the actual friction is much higher than that predicted by Eq 5a and 5b (see the Appendix to this article for additional information).

On portions of the apparent contact area, there is actually clearance between the workpiece and the die (Fig. 10). When lubrication is provided, some lubricant is dragged between the two surfaces by the moving workpiece. As the value of $\eta v_f/(R_f\sigma_0)$ increases, more of the voids between the surfaces are filled with lubricant. In addition, more of the pressure load on the workpiece is transmitted through the lubricant, and less load acts to smooth the asperities on th workpiece. Subsequently there is less actual metal-to-metal contact and friction drops. At a sufficiently high value of $\eta v_f/(R_f\sigma_0)$ or beyond this critical value, enough lubricant is dragged into the interface to cause complete separation of the two bodies. No metal-to-metal contact occurs, and full hydrodynamic lubrication with low friction exists. A further increase in $\eta v_f/(R_f\sigma_0)$ will cause a monotonic increase in the friction value and in the thickness of the lubrication layer.

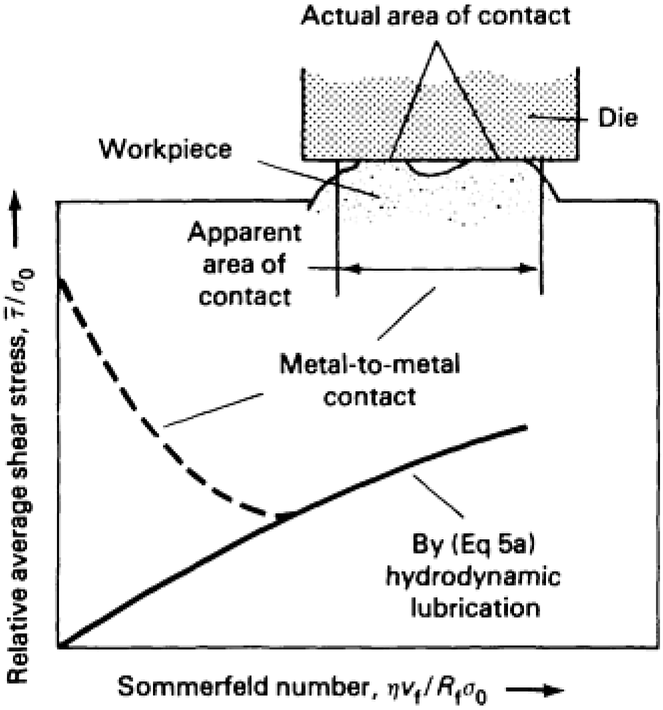


Fig. 10 Effect of Sommerfeld number on relative shear stress, friction, and lubrication characteristics

The onset of hydrodynamic lubrication is distinctly observable. Nevertheless, even when hydrodynamic lubrication prevails, a few occasional high spots on the workpiece will contact the die.

If one assumes an arbitrary value for the film thickness $(\epsilon/R_f)_{cr}$ at the exit as the critical onset value at which hydrodynamic lubrication commences, then by Eq 5aand 5b, the critical Sommerfeld number for the onset of hydrodynamic lubrication is:

$$S_{cr} = \left[\frac{\eta v_f}{\sigma_0 R_f}\right]_{cr} = \frac{f(\alpha)\,(\epsilon/R_f)^2_{cr}}{\frac{\cot\alpha\,\cos^2\alpha}{3}\,[1-(R_f/R_0)^3] + L/R_f} \quad \text{(Eq 6)}$$

from which the critical speed is readily determined. The critical Sommerfeld number decreases with an increase in the bearing length (L/R_f) or with a decrease in the die angle (α) or in the critical film thickness $(\epsilon/R_f)_{cr}$.

More detailed information on the phenomenon of hydrodynamic lubrication during flow through converging dies can be found in Ref 14.

Modeling of Strip Rolling

The following description of rolling is given in Ref 14:

> The main purpose of strip rolling, where the width of the strip is much greater than its thickness, is to reduce the latter. In the following discussions, both rolls have the same radius R_O and surface conditions, are equally powered, and run at the same velocity. They transfer energy to the strip through the friction between the two bodies [Fig. 11]. Under regular conditions, the

strip moves slower than the roll at the entrance and faster than the roll at the exit, with a neutral point in between at which both speeds are equal. The friction force acting between the entrance and the neutral point advances the strip between the rolls, while the friction acting on the exit side from the neutral point opposes the rolling action.

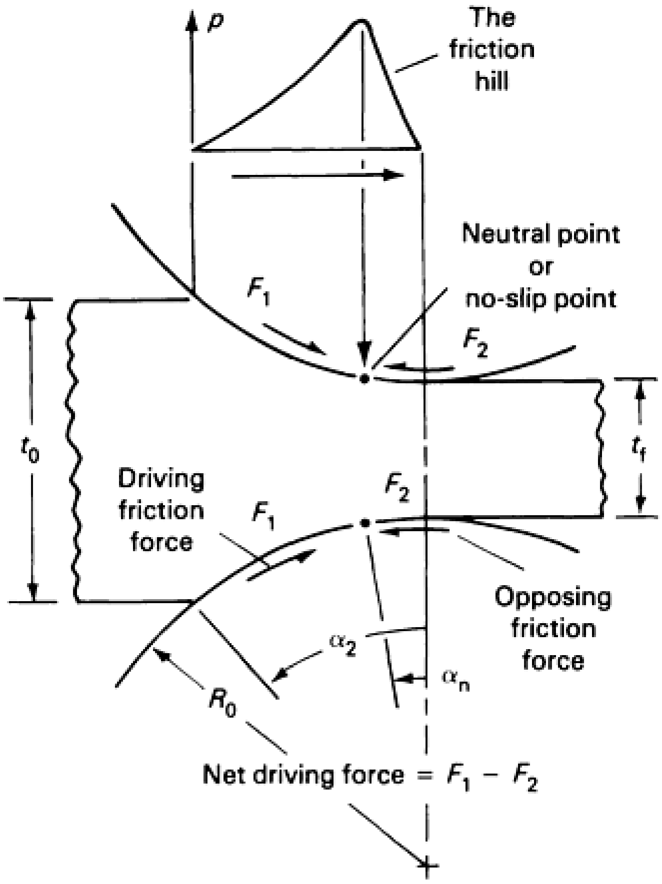


Fig. 11 Positive and negative friction directions during strip rolling

As reduction increases, the position of the neutral point approaches the exit. When the maximum possible reduction is attempted, the neutral point reaches the exit. This situation is very unstable. Any slight increase in reduction or drop in friction will cause the strip to stop moving and the rolls to start skidding over the strip. The distribution of the pressure between the rolls and the strip is described by the "friction hill" in Fig. 11. The peak of the friction hill occurs at the neutral point.

Hydrodynamic Lubrication. Lubricant is applied at the entrance side of each pair of rolls (stand). At the entrance side, the layers of lubricant that are in contact with the rolls or with the strip adhere to their respective metal surfaces and move inward toward the exit. An inlet entry zone, which is shaped as a wedge, also forms. The outer layers of this wedge move inward and a return flow of lubricant (in the form of an eddy current) occurs between the surface layers (Fig. 12a). At low rolling speeds, the entry zone wedge is negligibly small. With increasing rolling speeds (or increasing liquid viscosity or increasing values of the Sommerfeld number) the wedge increases both in thickness and depth. The Sommerfeld number is:

$$S = \frac{\text{Viscosity} \cdot \text{Strip exit velocity}}{\sigma_0 \cdot t_f} \quad \textbf{(Eq 7)}$$

With increasing values of *S*, the point where metal-to-metal contact between the strip and the rolls is established moves further toward the exit.

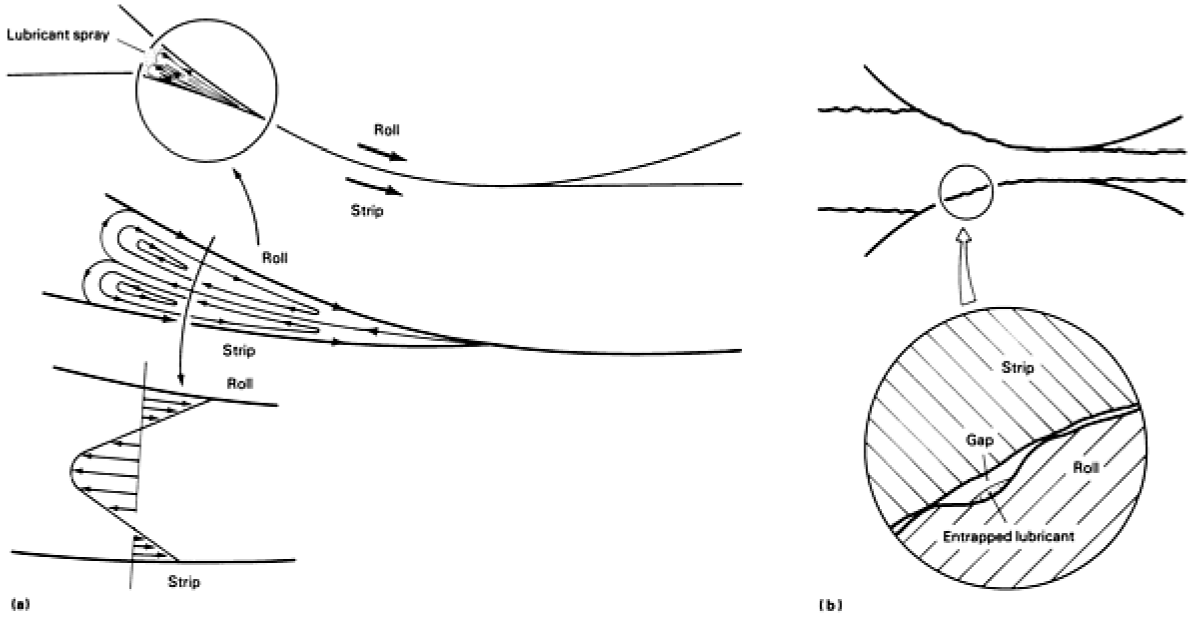


Fig. 12 The lubrication entry zone (a) and surface irregularities (b) associated with strip rolling

The surfaces of both rolls and strip are not perfectly smooth surfaces in that they contain irregularities in the form of peaks (or crests) and valleys (cavities) (Fig. 12b). Some lubricant passes from the entrance to exit side through the labyrinth of channels created by these irregularities. As the strip is deformed, the crests are flattened and the entrapped lubricant is pressurized in the diminishing volume of the cavities. At slow rolling speeds, the excess lubricant in the diminishing cavity space is squeezed to flow back into the entry zone. At higher rolling speed, the escape of excess fluid from the diminishing gap of the labyrinth between the rolls and the strip is relatively slower. The entrapped lubricant is then pressurized and causes partial separation between the rolls and the strip. At low speed, pressure is transmitted from the rolls to the strip through metal-to-metal contact. At higher speeds, more of the pressure is transmitted through the entry zone and through the entrapped lubricant. As the pressure transmitted through metal-to-metal contact is reduced with higher speeds, the friction is decreased. Eventually, at high speeds, no metal-to-metal contact exists.

When conditions for complete hydrodynamic lubrication are reached (Fig. 13), friction is at its minimum value. Therefore, with increased speeds, high shear rates are created in the liquid where the shears stress is proportional to the shear strain rate to the second power, and friction rises at a mild rate. Friction values are much lower when hydrodynamic lubrication prevails than when metal-to-metal contact is prevalent.

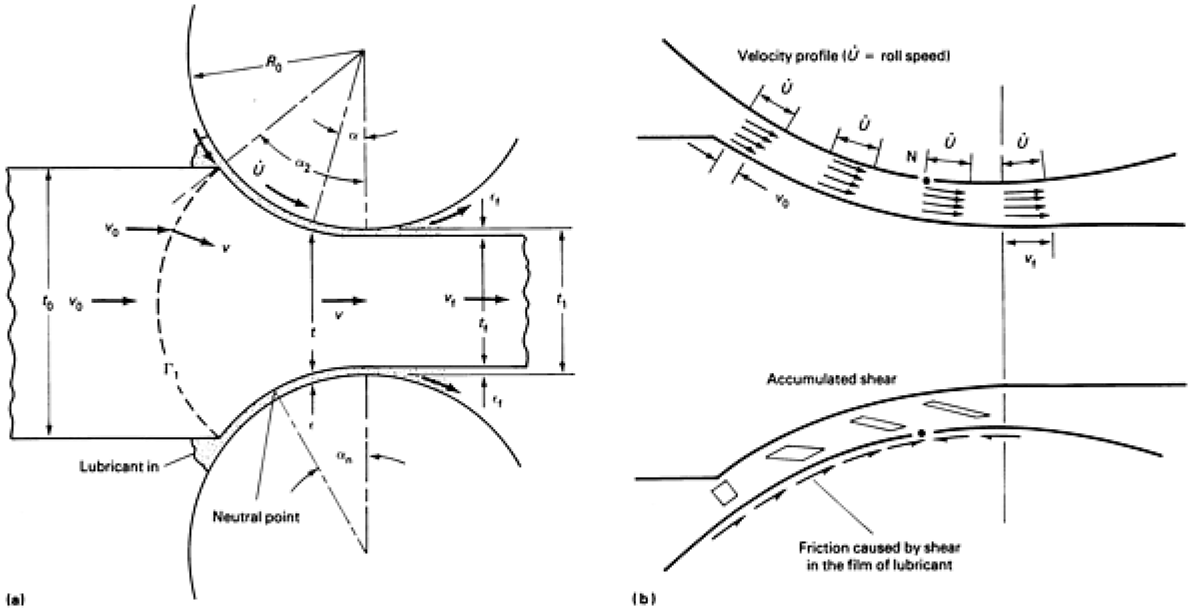


Fig. 13 Hydrodynamic lubrication during strip rolling. (a) Overall schematic. (b) Shear in the lubricant film

With increasing speed the friction hill effect is also reduced, and roll separation force and roll flattening become less pronounced. With reduced roll separation force, the elastic stretching of the mill as well as roll bending and roll flattening are reduced, causing the gap between the rolls to decrease. With the increasing thickness of the lubricant film and reduced mill flexing, the thickness of the emerging strip reduces. The actual gap between the rolls is larger than the thickness of the strip by twice the film thickness. Because friction drag decreases with increasing rolling speed, the neutral point approaches the exit and forward slip is reduced.

Two critical points may now be reached. The first is skidding due to insufficient friction drag while the second is the establishment of hydrodynamic lubrication. If the strip is thin enough, hydrodynamic lubrication is reached first and skidding will not develop. When speed continues to increase after hydrodynamic film lubrication is established, friction drag and forward slip, which have already reached their respective minimum points, begin to increase (see lines 1 to 4 in Fig. 14). For increasing values of strip thickness (lines 2, 3, and 4 in Fig. 14), the forward slip decreases and the critical value of roll speed at which the neutral angle and forward slip reach a minimum is increased. The values of the neutral angle and forward slip at which hydrodynamic lubrication begins increase with increasing thickness values.

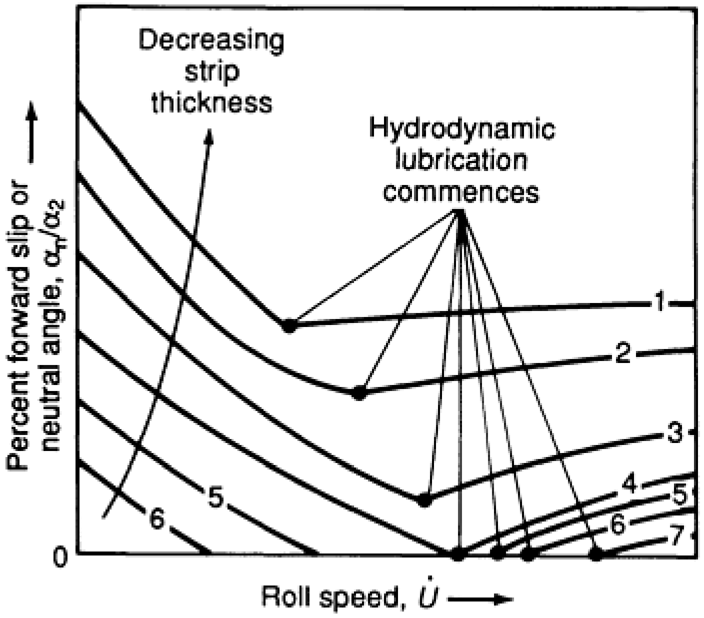


Fig. 14 Forward slip and position of the neutral point versus rolling velocity. See text for explanation of numbered curves.

At a critically high strip-thickness value, hydrodynamic lubrication commences when forward slip and the value of the neutral angle are zero (line 4 in Fig. 14). When the thickness of the strip is above the stated critical value, for lines 5 to 7, friction drops to below the minimum required for rolling before hydrodynamic lubrication can be established. Skidding then commences at the critical roll speed. It can be observed that at higher values of speed, referred to here as the minimum required speed, rolling with hydrodynamic lubrication may be reestablished. It should also be noted that the minimum required speed can be reduced by the use of lubricants of higher viscosity.

The following variations in the friction hill are expected with increasing speed:

- The peak of the friction hill gets lower and, together with the neutral point, shifts closer to the exit
- The entry zone expands and the corner of the liquid wedge moves further away from the entrance and closer to the exit. The meaningful slope of the friction hill on the entrance side starts at the area around the tip of the entry zone
- The roll separation force gets lower
- Roll bending/flattening and mill stretching are reduced
- Strip thickness is reduced
- For relatively thick strip (or low dry-friction value, m_o), a critical roll speed of the first kind may be reached. When this occurs, friction becomes so low (below the critical value required) that the neutral point is at the exit and skidding will commence (lines 5 and 6). At the moment that skidding begins, the strip stops
- For relatively thin strip another critical speed (the critical speed of the second kind) can be reached, even before skidding commences. When this occurs, the point of the entry zone will reach the exit and hydrodynamic lubrication will commence. A further increase in speed will cause a thicker lubricating film to separate the rolls from the strip and make the rolling more stable (see lines 1, 2, and 3 in Fig. 14)
- For relatively thick strip, even when a critical speed of the first kind commences first, a further increase in speed may ultimately produce a critical speed of the second kind and reestablish hydrodynamic lubrication (see lines 4 to 7 in Fig. 14)

Changes in the entry zone, the friction hill, the neutral point, and strip thickness as a function of speed for thick and thin strip are shown schematically in Fig. 15.

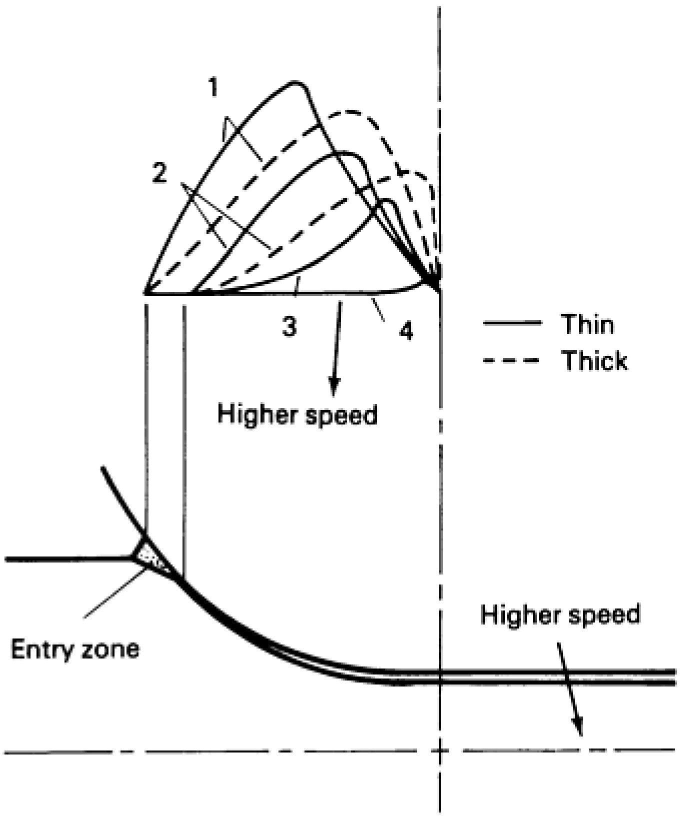


Fig. 15 Variations in the "friction hill" versus speed. The parameters *m, R_O*, σ_{xb}, σ_{xf}, and percent reduction are constant.

Measurement of Friction

Ring Forging. Figure 16 shows the deformation characteristics of two identical rings reduced in height incrementally by the same amount (ΔT_0). One ring is lubricated well and is pressed with low friction, while the unlubricated ring is pressed with high friction. The characteristic behavior is drastically different in the two cases. In the low-friction operation, the inner radius (R_i) of the ring increases in size, while in the high-friction operation, the inner radius of the ring decreases in size.

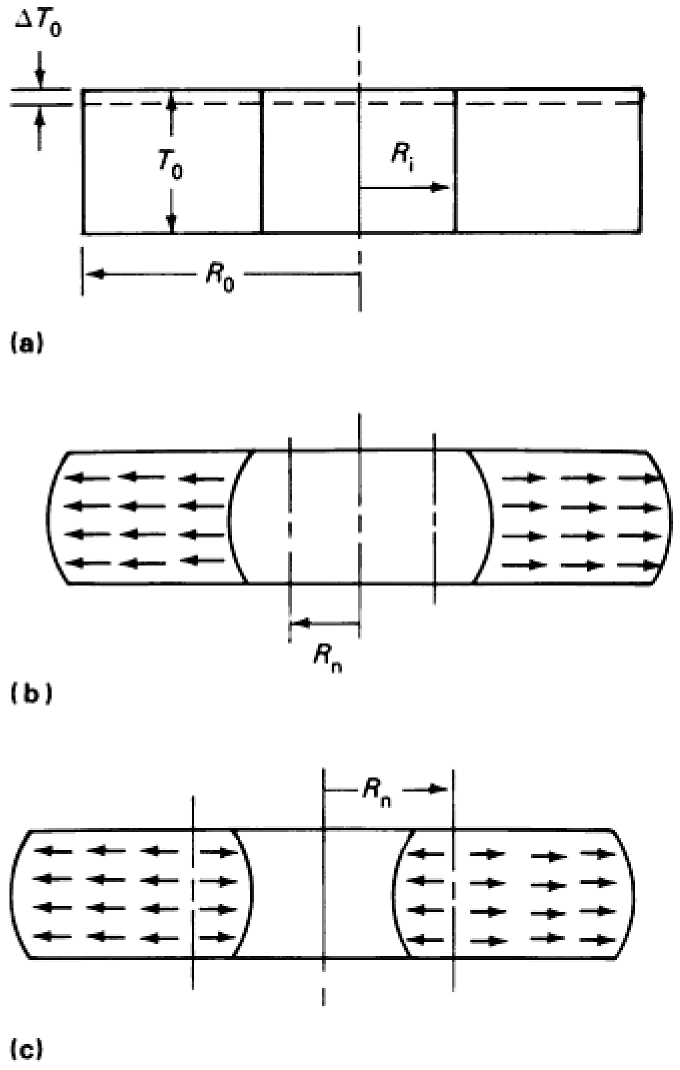


Fig. 16 Deformation behavior during ring forging. (a) Original ring. (b) After deformation with low friction. (c) After deformation with high friction

The bulge direction changes accordingly, minimizing relative sliding of the workpiece over the platens. The outer radius (R_o) expands in both cases.

In describing the flow pattern, consider an imaginary cylinder of radius $R = R_n$, where R_n is the neutral radius. In the high-friction ring, all points with radius $R > R_n$ move outward and their radial positions increase while all points with radial position $R < R_n$ are moving inward. Points on the radius $R = R_n$ are stationary, and hence this radius is called the neutral radius. In the high-friction ring of Fig. 16, the neutral radius resides in the ring itself, so that the neutral radius is larger than the inner radius ($R_n > R_i$). On the low-friction ring, the neutral radius is smaller than the inner radius. Therefore, all points on the low-friction ring are in a radial position $R > R_n$. The increase in outer radius for higher friction is smaller than that for low friction.

When $R_n \leq R_i$, the position of the neutral radius (R_n) can be determined as follows (Ref 15):

$$\left(\frac{R_n}{R_o}\right)^2 = \frac{\sqrt{3}}{2}\,\frac{1-(R_i/R_o)^4\,x^2}{\sqrt{x(x-1)\,[1-(R_i/R_o)^4 x]}}$$

where **(Eq 8a)**

$$x = \left\{\frac{R_o}{R_i}\exp\left[-m\,\frac{R_o}{T}\left(1-\frac{R_i}{R_o}\right)\right]\right\}^2$$

When $R_i \le R_n \le R_o$ the position of the neutral radius is found by successive approximations from:

$$2m\,\frac{R_o}{T}\left(1+\frac{R_i}{R_o}-2\,\frac{R_n}{R_o}\right) + \ln\left|\left(\frac{R_i}{R_o}\right)^2 \frac{(R_n/R_o)^2+\sqrt{3+(R_n/R_o)^4}}{(R_n/R_o)^2+\sqrt{3\,(R_i/R_o)^4+(R_n/R_o)^4}}\right| = 0$$

(Eq 8b)

When a ring is forged and all the geometrical parameters, including the value of R_n, are determined experimentally, then the friction factor *m* is calculated, using Eq 8a and 8b. References 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 provide additional information on ring test friction analysis, as follows:

- *Analysis.* Slug or slab free body equilibrium approach: M. Kunogi (Ref 16), 1956. Upper bound, the parallel velocity field: I. Tarnovski *et al.* (Ref 17), 1959; H. Kudo (Ref 18), 1961; B. Avitzur (Ref 29), 1964. Upper bound with bulge: B. Avitzur (Ref 19), 1969. Limit analysis (that is, upper and lower bounds): B. Avitzur and F.R. Sauerwine (Ref 20, 21), 1978
- *Experimental.* Use of feeler pins in disk forging: G.T. Van Rooyen and W.A. Backofen (Ref 22), 1960. Calibration curves: A.T. Male and M.G. Cockroft (Ref 23), 1964
- *Combined.* Analytically supported calibration curves: A.T. Male *et al.* (Ref 24, 25), 1970

Flow through Conical Converging Dies. Equation 4 and Fig. 5 show that during flow through conical converging dies, the forces required are dependent on the die cone angle. There always exists an optimal cone angle that requires the minimum force. With a cone angle smaller than the optimal cone angle, the drawing or extrusion force is high because the length of contact between the die and material is high, causing excessive friction losses. With a die of an angle larger than the optimal, the distortion is excessive, causing high forces.

When Eq 4 is differentiated with respect to α and the derivative is equated to zero, the resulting equations express implicitly the relation between the semicone angle of the die, which minimizes the drawing or extrusion stresses and the other variables.

Experimentally the optimal die angle can be determined by following the procedure offered by Wistreich (Ref 26). Several dies of identical exit diameter are used, each of different entry cone angle. Rods of identical material, treatment, diameter, surface condition, and lubrication are formed through the dies. The forming force (or pressure) is measured. A plot (Fig. 5) of the drawing or extrusion stress is obtained as a function of cone angle, assuming that reduction, friction, and flow stress are all constant. The point of minimum stress indicates on the abscissa the value of the optimal die angle (α_{opt}) for the experienced reduction and friction. When a second batch of rods of different initial diameter is drawn (or extruded), different values for the optimal die angle are obtained. The friction that prevails during each run of the experiment can be computed as follows:

$$m = 2\left\{\frac{1 - \alpha_{opt}\cot\alpha_{opt}}{\ln(R_0/R_f)} + \sqrt{3}(\sin\alpha_{opt}) \times \left[\sqrt{1 - \left(\frac{11}{12}\right)\sin^2\alpha_{opt}} - (\cos\alpha_{opt})f(\alpha_{opt})\right]\right\} \quad \textbf{(Eq 9)}$$

Strip Rolling. Avitzur (Ref 27, 28, 29) has developed an experimental procedure for the determination of the coefficient of friction during strip rolling. The expressions derived for these studies can be used to determine friction values when skidding commences.

With the assumption of a constant shear factor (see Eq 2 in this article), Eq 22 from Ref 29 reads:

$$m = \sqrt{\frac{t_f}{R_O}} \cdot \frac{\ln\left(\frac{t_0}{t_f}\right) + \frac{1}{4}\sqrt{\frac{t_f}{R_0}}\sqrt{\frac{t_0}{t_f} - 1} + \sigma}{\tan^{-1}\sqrt{\frac{t_0}{t_f} - 1}} \quad \textbf{(Eq 10)}$$

where

$$\sigma = \frac{\sigma_{xb} - \sigma_{xf}}{\frac{2}{\sqrt{3}}\sigma_0}$$

When skidding occurs, the following have to be measured: relative front $\sigma_{xf}/(2\sigma_0/\sqrt{3})$ and back $\sigma_{xb}/(2\sigma_0/\sqrt{3})$ tensions, roll radius (R_O), and incoming (t_0) and emerging (t_f) strip thicknesses. The value of friction is computed then by Eq 10. When no front or back tensions are applied, skidding can be induced only with small roll diameter. When $\sigma_{xf} = \sigma_{xb} = 0$, Eq 10 reduces to:

$$m = \sqrt{\frac{t_f}{R_O}}\,\frac{\ln(t_0/t_f) + (1/4)\sqrt{t_f/R_O}\sqrt{t_0/t_f - 1}}{\tan^{-1}\sqrt{t_0/t_f - 1}} \quad \textbf{(Eq 11)}$$

When $\sigma_{xf} = \sigma_{xb} = 0$, only the geometrical parameters t_0, t_f, and R_O have to be measured and a rolling mill with no instrumentation can be used for the evaluation of the coefficient of friction.

When $\sigma_{xf} = \sigma_{xb} = 0$ and skidding occurs, the following relation is proposed:

$$\alpha_2 = 2\mu \quad \textbf{(Eq 12)}$$

where α_2 is the contact angle. By replacing α_2 by strip thickness and roll radius, the following is written:

$$\mu = \left(\frac{1}{2}\right)\sqrt{t_f/R_O}.\sqrt{(t_o/t_f) - 1} \quad \textbf{(Eq 13)}$$

In a study detailed in Ref 27, strips of identical thickness were rolled with varying gaps between the rolls. The strip was pushed between the rolls to start the operation. If the full length of the strip passed the rolls, the gap was narrowed and

another strip fed in. This procedure went on until skidding started. When skidding started, the original and final thickness of that strip and the radius of the rolls were recorded. Another batch of strips with smaller original thickness, but the same annealed and surface conditions, was then rolled with the same lubricant. The previous procedure was repeated until skidding started. Several batches of various original strip thickness were tested. The friction values were computed using Eq 11 for each batch.

Appendix: The Wave Model

The model presented in Ref 30 has provided a quantitative description of Coulomb's (Ref 31) and Amonton's (Ref 32) characterizations of friction for moderate loads. This model is based on the "wave model" concept of the mobility of the asperities.

In his *History of Tribology* (Ref 33), Dowson explains the early recognition of the cause-and-effect relationships between surface irregularities and the resistance to sliding. Dowson summarizes Coulomb's ideas on the combined effects of the then-prevailing thoughts on cohesion and surface irregularities. Coulomb recognized that irregularities caused resistance to sliding, and sketched the opposing asperities as enmeshed fibers folding over and filling the voids previously occupied by opposing surfaces as sliding takes place. However, as Dowson points out, the energy dissipation was not then recognized because the science of thermodynamics had not yet been born. The theory of folding asperities would later be replaced by the concept of the mobile ridge.

Leslie observed that asperities on the harder surface (called "wedges" here) push on the opposing ridges of the softer surface. Unlike the popular theory of those days that a climbing motion occurred, ignoring the inevitable downhill companion motion, Leslie suggested that the softer ridge was pushed into and under the surface, producing a perpetual uprising ahead of the wedge of the harder surface and resulting in an endless climb (Ref 34). Evidence for the mobility can be seen in Fig. 17, which is reproduced from Ref 35. Simulations of plastic deformation of the asperities by the slip line and the upper bound techniques are presented in Ref 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49.

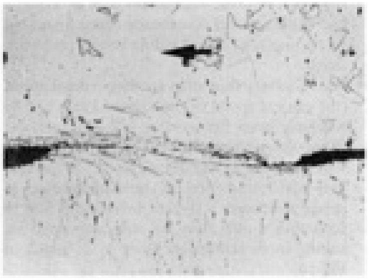


Fig. 17 Photomicrograph showing the mobility of asperities and the thin layer of deformations. Source: Ref 35

Presently, the model of the mobile ridge and the trapped lubricant on the interface between two sliding surfaces is examined quantitatively to provide the characteristics shown in Fig. 4(a) and of the Stribeck curve (Ref 50) shown in Fig. 4(b). To model the interaction between the asperities, the photomicrograph shown in Fig. 17 can be replaced by the illustration of wedge flow in Fig. 18. The tip of the wedge on the surface of the die indents into the surface of the workpiece, producing a ridge. Sliding between the two surfaces is made possible by the ridge being suppressed under the wedge when it passes under it. The material ahead of the contact is pushed upward, and the ridge experiences a wave motion. The lubricant trapped between the stationary wedge and the moving ridge is being forced to move with the solid that it contacts, and thus experiences an eddy flow. Figure 19 shows theengineering model for the flow depicted in Fig. 18. Here the deformation region of the ridge is described by three rigid triangles, each moving with respect to each other, through a sliding motion along their interfaces. In Ref 45, the triangles move in a linear motion, while in later studies (Ref 47), they move in the more general rigid-body motion (rotational motion).

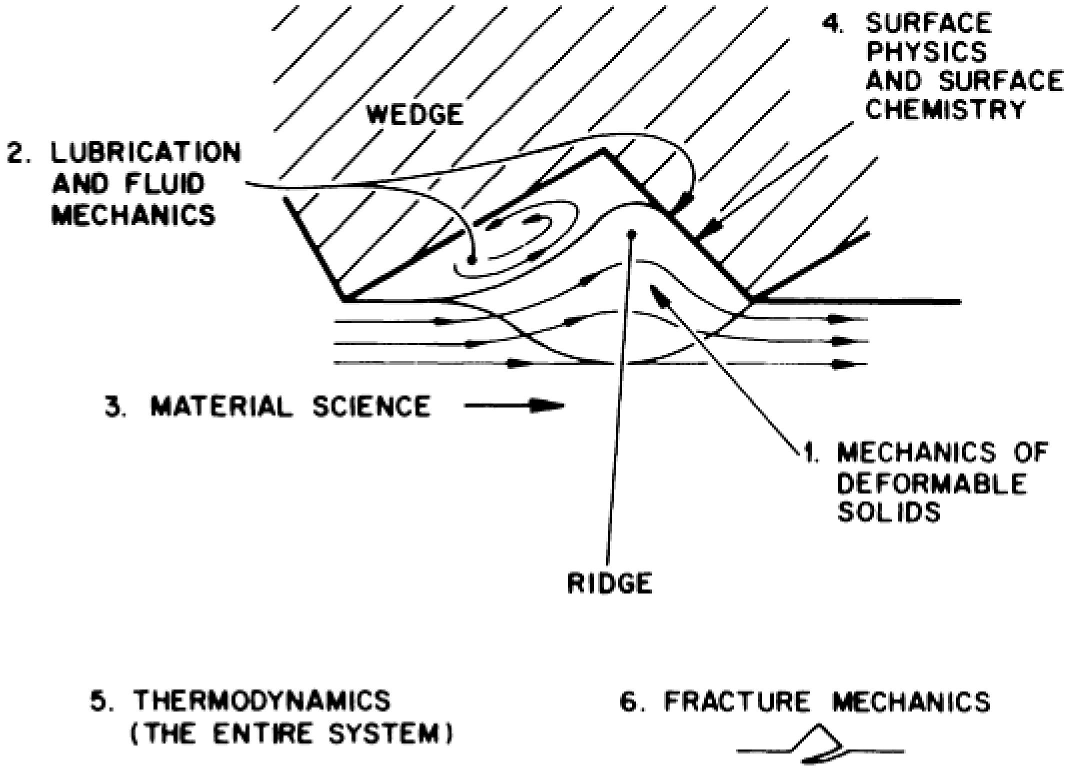


Fig. 18 Disciplines affecting wedge flow

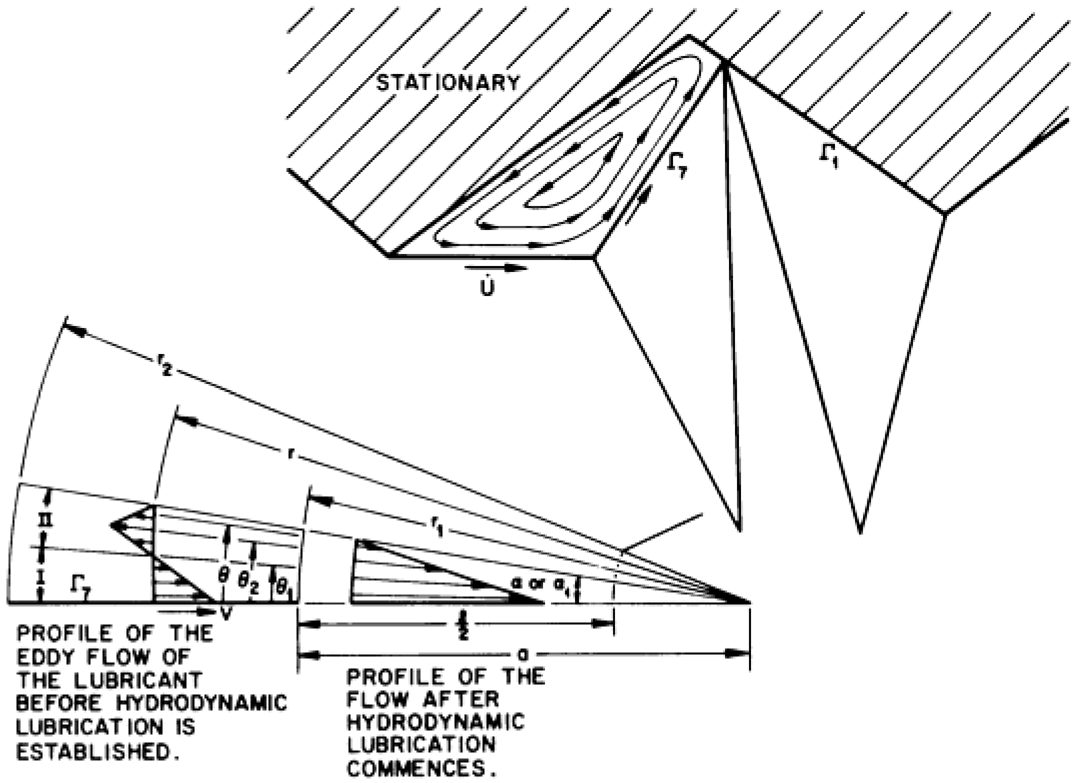


Fig. 19 Engineering model for wedge flow showing velocity field in the lubricant

The Characteristics of Friction

The model shown in Fig. 19 was used in Ref 44 and 45 to obtain friction characteristic by the upper-bound approach. The relationship between the friction force ($F/\sigma_0\ell$) and relative pressure ($p/\sigma_0\ell$) with different input parameters is given in Fig. 20 and 21. In Fig. 20, the characteristics of the global friction at the ordinate are described as a function of the local friction factor (m_0 as the input parameter, while the angle of the asperity (α_1) is constant at 5°. In Fig. 21, the angle of the asperity (α_1) is the parameter, while the local friction is constant at m_0 = 0.2. Global friction increases first linearly with increasing pressure and then tapers off asymptotically to a constant value. Global friction also increases with increasing values of local friction and with increasing asperity inclination.

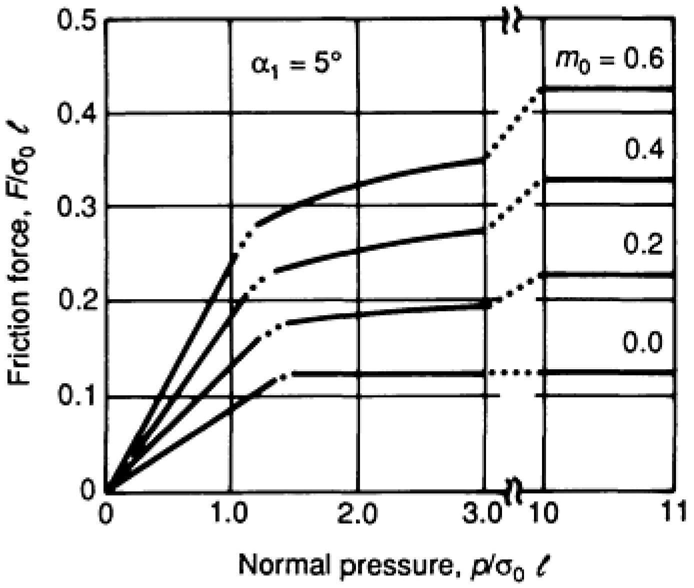


Fig. 20 The relation between the friction force and normal pressure with friction factor (m_0) as a parameter

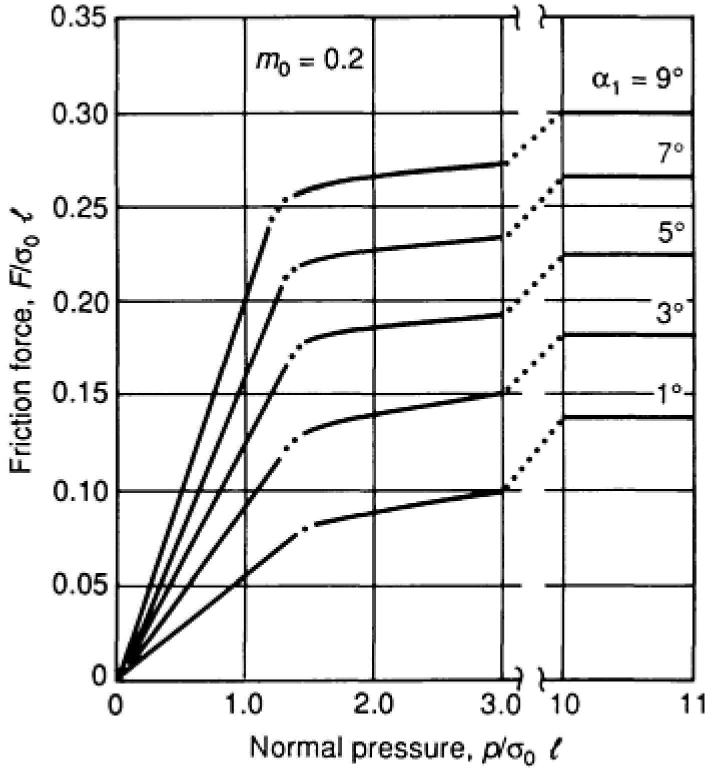


Fig. 21 The relation between the friction force and normal pressure with wedge angle (α_1) as a parameter

One observation worth nothing is that even when the local friction (m_0) is assumed to be zero, the global friction (m) is not zero. Resistance to sliding exists due to the deformation of the asperities at the surface. Another observation is that the global friction factor never reaches the value of $m = 1$. For high values of m_0 and/or of α_1, the wave model is replaced by the ridge shear failure model shown in Fig. 3(a).

The range of the resulting changes in the power consumed through the mobility of the ridge, and through shear losses in the trapped lubricant between the asperities due to the eddy flow, is wider than shown in Fig. 4(b) as demonstrated in Ref 51, 52, 53, 54. This complexity is complexity is evident from the calculated value of the global friction factor (m) as presented in Ref 51, and Fig. 22 and 23. In Fig. 22, the abscissa is the Sommerfeld number (S), where the dimensionless Sommerfeld number combines the velocity of sliding (v_0), the viscosity of the lubricant (η), the strength of the material (σ_0), and the length of the asperity (ℓ_0), in the form of $S = v_0\eta/(\sigma_0\ell_0)$. The ordinate is the global friction factor (m) and the parameter is the normal load (p) on the interface between the two sliding bodies. The local friction factor is $m_0 = 0.6$ and the inclination of the asperity is $\alpha_1 = 1$ °. For the lower load values ($p = 2$), the characteristic behavior shown in Fig. 4(b) is observed. The static friction factor value of m is highest when no sliding occurs. With increasing sliding speeds friction drops. Higher pressure values produce higher resistance to sliding. For higher pressures, the ridge is higher, and therefore the thickness of the film of the trapped lubricant is thinner. Furthermore, increases in the Sommerfeld number are not as effective in reducing the height of the ridge. Thus, for high pressures, the lubricant film remains thin even with increasing Sommerfeld number.

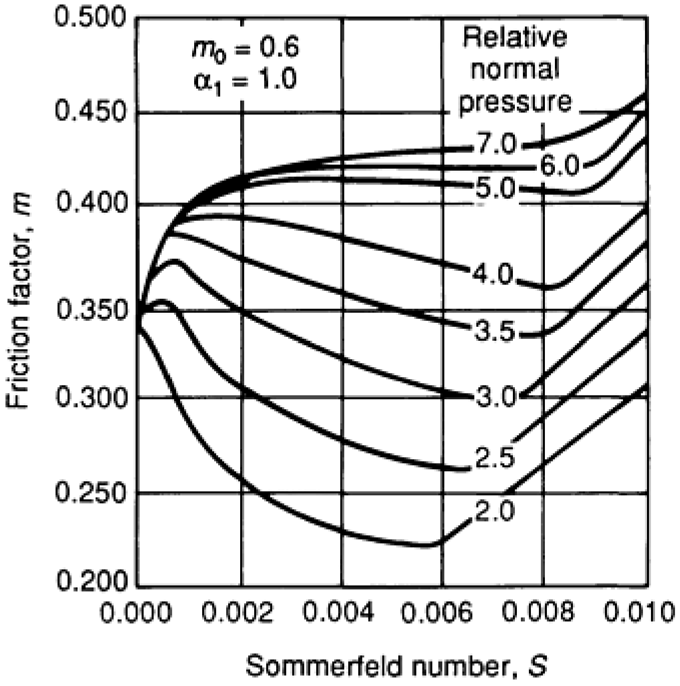


Fig. 22 Global friction versus Sommerfeld number at high pressure

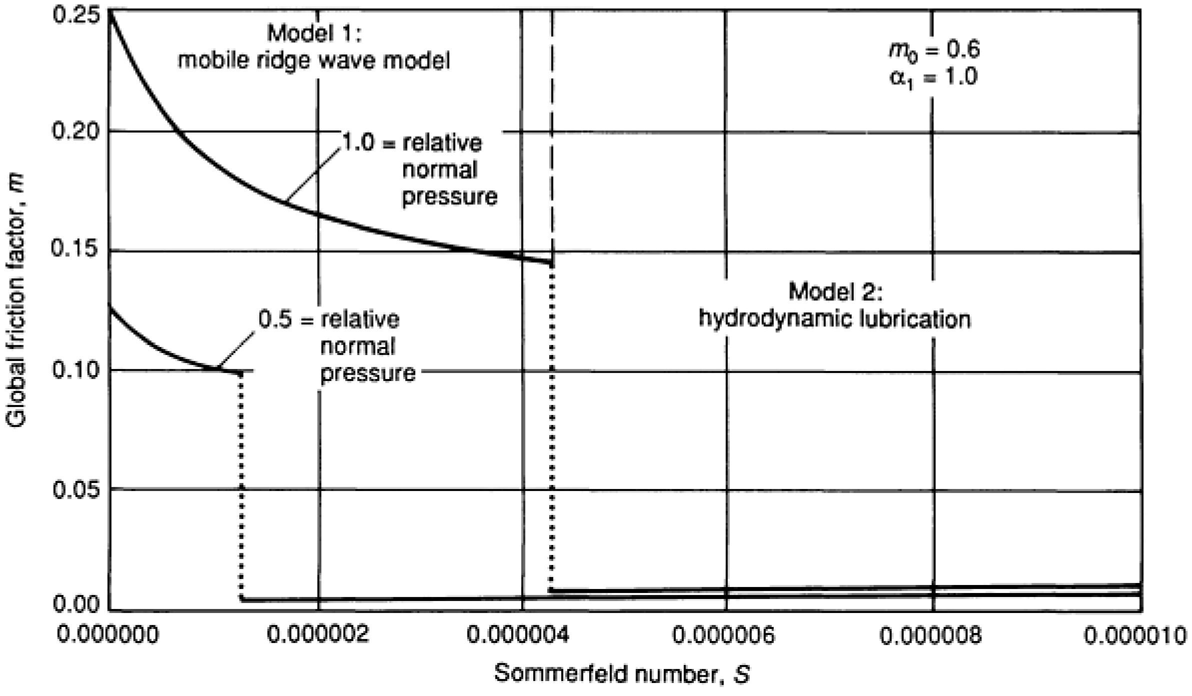


Fig. 23 Transition from model 1 (the mobile ridge) to model 2 (the hydrodynamic lubrication) friction mode

The two inserts in the lower left-hand corner of Fig. 19 display boundary lubrication (by model 1 with ridge mobility) and hydrodynamic lubrication (by model 2 with full film separation between the tool and the workpiece). The transition from model 1 (mobile ridge) to model 2 (hydrodynamic lubrication) is illustrated in Fig. 23. For small angles (α_1 = 0.1) and low pressure (p/σ_0 = 0.5), a gradual decrease in the global friction value (*m*) with increasing Sommerfeld number (*S*) is observed. When the critical value of the Sommerfeld number is reached and hydrodynamic lubrication commences, the friction drops abruptly to $\leq$2% of the static value.

The drastic friction drop presented by the wave model has also been observed in other treatments of the onset of hydrodynamic lubrication during wire drawing and rolling (Ref 12). However, in the present study, it is worth noting that the abrupt drop in friction when hydrodynamic lubrication commences is expected only if all asperities are identical. Asperities come in all sizes and shapes, and the critical Sommerfeld number for each occurs at a different value. Thus, the actual decrease in *m* is gradual. The increase in *m* with further increase in *S* is very slight.

Additional information on the wave model, including comparisons of some of the studies/results described in this article, can be found in Ref 54, 55, 56.

References

1. F.P. Bowden and D. Tabor, *The Friction and Lubrication of Solids,* Oxford University Press, Oxford, Part I, 1954; Part II, 1964
2. N. Bay, Analysis of Tool-Workpiece Interface Stresses in Metal Forming, *Proceedings of NAMRC 14* (University of Minnesota, Minneapolis), Society of Manufacturing Engineers, 1986, p 388-393
3. B. Avitzur, The Use of the Personal Computer for Simulation of the Process of Wire Drawing and Extrusion in an Interactive, User Friendly Mode, *Proceedings of the Wire 89 International Conference* (Atlanta, Georgia), 5-10 Nov 1989, p 396-403
4. M.J. Hillier, A Hydrodynamic Lubrication in Wire Drawing, *Int. J. Prod. Res.,* Vol 5, 1967, p 171
5. A.F. Gerds and F.W. Boulger, Rod, Wire and Tube Drawing, DMIC Report 226, *Metal Deformation Processing,* Vol II, Defense Metals Information Center, Battelle Memorial Institute, 7 July 1966, p 78-124
6. *Recent Progress in Metal Working,* American Elsevier, 1964

7. G.H. Tattersall, "Theory of Hydrodynamic Lubrication in Wire Drawing," Report MW/D/46/59, British Iron and Steel Research Association

8. D.G. Christopherson and H. Naylor, Promotion of Fluid Lubrication in the Wire Drawing, *Proc. Inst. Mech. Eng.,* Vol 169, 1955, p 643-653

9. V.F. Moseev and A.A. Korostilin, New Method of Feeding Lubricant to the Deformed Zone in Wire Drawing (in English), *Stal',* Vol 3, Mar 1962, p 237-239

10. L.H. Butler, A Method for Continuous Wire Drawing Aided by Externally Generated Hydrostatic Oil Pressure, *J. Inst. Met.,* Vol 93, 1964-1965, p 123-1

11. H.D. Pugh, Recent Developments in Cold Forming, *Bulleid Memorial Lectures* (University of Nottingham, England), Vol IIIA and IIIB, 1965

12. B. Avitzur, *Metal Forming: Process and Analysis,* McGraw-Hill, 1968, Krieger, revised 1979

13. A. Bobrowsky, "Pressure Box," U.S. patent 3,417,589, 24 Dec 1968

14. B. Avitzur, *Handbook of Metalforming Processes,* John Wiley, 1983

15. B. Avitzur, Forging of Hollow Discs, *Isr. J. Technol.,* Vol 2 (No. 3), 1964, p 295-304

16. M. Kunogi, *Rep. Sci. and Res. Inst.,* Vol 30 (No. 63), 1954 (in Japanese with English summary); M. Kunogi, A New Method of Cold Extrusion, *J. Sci. Res. Inst.,* Vol 50 (No. 1437), 1956, p. 215-246

17. I. Tarnovski, A.A. Pozdeev, and O.A. Ganago, *Deformation and Forces in Metal Forming,* National Science and Technology Co., Machine Building Literature, Moscow (in Russian), 1959

18. H. Kudo, Some Analytical and Experimental Studies of Axisymmetric Cold Forging and Extrusion, Parts I & II, *Int. J. Mech. Sci.,* Vol 2, 1960, p 102-127, and Vol 3, 1961, p 91-117

19. B. Avitzur, "Bulge in Hollow Disc Forging," Report of the Institute for Metal Forming, Lehigh University, Aug 1969, and Technical Report AFML-TR-69-261, Air Force Materials Laboratory, Air Force Systems Command, Nov 1969

20. B. Avitzur and F.R. Sauerwine, Limit Analysis of Hollow Disc Forging, Part I: Upper Bound, *J. Eng. Ind. (Trans. ASME),* Vol 100 (No. 3), Aug 1978, p 340-346

21. F.R. Sauerwine and B. Avitzur, Limit Analysis of Hollow Disc Forging, Part 2: Lower Bound, *J. Eng. Ind. (Trans. ASME),* Vol 100 (No. 3), Aug 1978, p 347-355

22. G.T. van Rooyen and W.A. Backofen, A Study for Interface Friction in Plastic Compression, *Int. J. Mech. Sci.,* Vol 1, 1960, p 1-27

23. A.T. Male and M.G. Cockroft, A Method for the Detonation of the Coefficient of Friction of Metals under Conditions of Bulk Plastic Deformation, *J. Inst. Met.,* Vol 93, 1964, p 38-46

24. V. Depierre and A.T. Male, The Validity of Mathematical Solutions for Determining Friction from the Ring Compression Test, *J. Lubr. Technol. (Trans. ASME),* Vol 92, 1970, p 389-395

25. V. Depierre, G. Saul, and A.T. Male, "The Relative Validity of Coefficient of Friction and Interface Friction Shear Factor for Use in Metal Deformation Studies," Technical Report AFML-TR-70-243, Air Force Materials Laboratory, Air Force Systems Command, Oct 1970

26. J.G. Wistreich, Investigation of the Mechanics of Wire Drawing, *Proc. Inst. Mech. Eng.,* Vol 169, 1955, p 654-670

27. B. Avitzur, Maximum Reduction in Cold Strip Rolling, *Proc. J. Mech. Eng. Ind.,* Vol 174 (No. 32), 1960, p 865-884

28. B. Avitzur, Power Analysis of Cold Strip Rolling, *J. Eng. Ind. (Trans. ASME),* Series B, Vol 85 (No. 1), Feb 1963, p 77-88

29. B. Avitzur, An Upper Bound Approach to Cold Strip Rolling, *J. Eng. Ind. (Trans. ASME),* Series B, Vol 86, 1964, p 31-48

30. B. Avitzur, *Wear,* Vol 126, 1988, p 227-249

31. A. Coulomb, Theorie des machines simples, en egant egard au frottement de leurs partres, et a la roideur des cordages, *Mém. Math. Phys.,* 1785, p 161-342

32. G. Amonton, Histoire de l'academie royale des sciences, *Mém. Math. Phys.,* 1699, p 206, and De la resistance causee dans les machines, *Mém. Acad. R., A,* 1706, p 257-282

33. D. Dowson, *History of Tribology,* Longman, London, 1979
34. J. Leslie, An Experimental Inquiry into the Nature and Propagation of Heat, *J. Newman,* No. 22, 1804, p 300, 302
35. M. Cocks, *Wear,* Vol 9, 1966, p 320-328
36. J.M. Challen, L.J. McLean, and P.L.B. Oxley, *Proc. R. Soc. (London) A,* Vol 394, 1984, p 161-181
37. N. Bay, Tool-Workpiece Interface Stresses in Cold Forward Extrusion, *Proceedings of the 1st International Conference on the Technology of Plasticity* (Tokyo), Japan Society for the Technology of Plasticity, 1984
38. T. Wanheim and T. Abildgaard, A Mechanism for Metallic Friction, *Proceedings of International Conference of Plastic Engineers* (Tokyo), Japan Society for the Technology of Plasticity, 1980
39. D. Tabor, *J. Lubr. Technol.,* Vol 103, 1981, p 169-179
40. H. Kudo, *Int. J. Mech. Sci.,* Vol 2, 1960, p 102-127; *Int. J. Mech. Sci.,* Vol 3, 1961, p 91-117
41. E. Rabinowicz, *Friction and Wear of Materials,* Wiley, 1966
42. C.M. Edward, *J. Mech. Eng. Sci.,* Vol 101, 1968, p 101-110
43. B. Avitzur and Y.D. Zhu, A Friction Model Based on the Upper Bound Approach to the Ridge and Sublayer Deformations Update, *Proceedings of the 13th NAMRC* (Berkeley, CA), Society of Manufacturing Engineers, May 1985, p 103-109
44. B. Avitzur, C.K. Huang, and Y.D. Zhu, *Wear,* Vol 95, 1984, p 59-77
45. B. Avitzur and Y. Nakamura, *Wear,* Vol 107, 1986, p 367-383
46. T. Wanheim, Friktion ved Hoje Flatetyk, Technical University of Denmark, 1969 (in Danish)
47. B. Avitzur and E. Kay, Surface-Layer Modeling of Friction Using Rotational Motion of Rigid Bodies, *Proceedings of the 3rd International Conference on the Technology of Plasticity* (Kyoto, Japan), Japan Society for the Technology of Plasticity, 1990, p 981-987
48. T. Abildgaard and T. Wanheim, An Investigation Into the Mechanisms of Abrasive Wear and Processing of Metals, *Proceedings of the 2nd Cairo University Mechanical Design and Production Conference,* Cairo University, Giza, 1982, p 521-529
49. H. Petryk, Slip-Line Field Solutions for Sliding Contact, *Proceedings of the International Mechanical Engineering Conference on Tribology: Friction, Lubrication, and Wear,* Institution of Mechanical Engineers, London, 1987, p 987-994
50. R. Stribeck, *Z. Ver. Deut. Ing.,* Vol 46 (No. 36), 1902, p 180
51. B. Avitzur, Boundary and Hydrodynamic Lubrication, *Wear,* Vol 139, 1990, p 49-76
52. H. Kudo, *Int. J. Mech. Sci.,* Vol 7, 1965, p 383
53. E. Felder, Interaction between Friction, Lubrication and Surface Roughness in Metal Working, *Surface Roughness Effects in Lubrication,* D. Dowson, C.M. Taylor, M. Godet, and D. Berthe, Ed., The Institute of Tribology, Leeds University, 1977, p 308-312
54. H. Kudo, S. Tanaka, K. Imamura, and K. Suzuki, *CIRP Ann.,* Vol 179G, 1976
55. N. Bay, T. Wanheim, and B. Avitzur, Modeling of Friction in Metal Forming, *Proceedings of NAMRC 15* (Bethlehem, PA), 27-29 May 1987, p 372-379
56. J.M. Challen and P.L.B. Oxley, *Wear,* Vol 100, 1984, p 171-193

Appendix: Static and Kinetic Friction Coefficients for Selected Materials

Peter J. Blau, Metals and Ceramics Division, Oak Ridge National Laboratory

THE DEFINITIONS for static and kinetic friction coefficients are given in the Glossary of this Handbook. The friction coefficient between solids sliding, or about to slide, over one another under the influence of a nonzero normal force is a function of several factors whose relative contributions vary on a case-by-case basis:

- Composition of the materials
- Surface finish of each solid
- Nature of the surrounding environment
- Force holding the solids in contact (load)
- Velocity of relative motion
- Nature of the relative motion (for example, unidirectional, back and forth, steady, variable, and so on)
- Nature of the contact (conforming versus nonconforming surfaces)
- Temperature of the interfacial region
- Prior sliding history of the surfaces
- Characteristics of the machine and fixtures in which the materials are affixed

No single source has generated a comprehensive list of friction coefficients for materials under identical testing conditions; therefore, nearly all existing handbooks rely on compilations of data produced under a variety of testing conditions. Readers should be aware of this shortcoming and use the values only as very approximate guides, unless their applications are exactly the same as those methods used in generating the data.

The five tables of friction coefficient values in this Appendix contain both static and kinetic friction coefficients. They are arranged by material type as follows:

- Table 1: metals on metals
- Table 2: ceramics on various materials
- Table 3: polymers on various materials
- Table 4: coatings on various materials
- Table 5: miscellaneous materials

It should be emphasized that the data in the tables are for unlubricated solids at room temperature and in ambient air. The reference list provided with each table lists both the sources of the data for the table and a brief description of the testing conditions used to generate these data, if such information was available in the reference. If accurate friction information is required for a specific application, the use of carefully simulated conditions or instrumentation of the actual machine should be conducted in lieu of using tabulated values because even a small change in contact conditions (for example, sliding speed or relative humidity for some materials) can result in a marked change in the measured or apparent friction coefficient.

Table 1 Friction coefficient data for metals sliding on metals

Metals tested in air at room temperature

Material		Test geometry [a]	Friction coefficient		Ref
Fixed specimen	**Moving specimen**		**Static**	**Kinetic**	
Ag	Ag	IS	0.50	. . .	1
	Au	IS	0.53	. . .	1
	Cu	IS	0.48	. . .	1
	Fe	IS	0.49	. . .	1
Al	Al	IS	0.57	. . .	1
	Ti	IS	0.54	. . .	1
Al, alloy 6061-T6	Al, alloy 6061-T6	FOF	0.42	0.34	2
	Cu	FOF	0.28	0.23	2
	Steel, 1032	FOF	0.35	0.25	2
	Ti-6Al-4V	FOF	0.34	0.29	2
Au	Ag	IS	0.53	. . .	1
	Au	IS	0.49	. . .	1
Brass, 60Cu-40Zn	Steel, tool	POR	. . .	0.24	3
Cd	Cd	IS	0.79	. . .	1
	Fe	IS	0.52	. . .	1
Co	Co	IS	0.56	. . .	1
	Cr	IS	0.41	. . .	1
Cr	Co	IS	0.41	. . .	1
	Cr	IS	0.46	. . .	1
Cu	Co	IS	0.44	. . .	1
	Cr	IS	0.46	. . .	1
	Cu	IS	0.55	. . .	1
	Fe	IS	0.50	. . .	1
	Ni	IS	0.49	. . .	1
	Zn	IS	0.56	. . .	1
Cu, OFHC	Steel, 4619	BOR	. . .	0.82	4
Fe	Co	IS	0.41	. . .	1
	Cr	IS	0.48	. . .	1
	Fe	IS	0.51	. . .	1
	Mg	IS	0.51	. . .	1
	Mo	IS	0.46	. . .	1
	Ti	IS	0.49	. . .	1
	W	IS	0.47	. . .	1
	Zn	IS	0.55	. . .	1
In	In	IS	1.46	. . .	1
Mg	Mg	IS	0.69	. . .	1
Mo	Fe	IS	0.46	. . .	1
	Mo	IS	0.44	. . .	1
Nb	Nb	IS	0.46	. . .	1
Ni	Cr	IS	0.59	. . .	1
	Ni	IS	0.50	. . .	1
	Pt	IS	0.64	. . .	1
Pb	Ag	IS	0.73	. . .	1
	Au	IS	0.61	. . .	1
	Co	IS	0.55	. . .	1
	Cr	IS	0.53	. . .	1
	Fe	IS	0.54	. . .	1
	Pb	IS	0.90	. . .	1
	Steel	SPOF	. . .	0.80	5
Pt	Ni	IS	0.64	. . .	1
	Pt	IS	0.55	. . .	1
Sn	Fe	IS	0.55	. . .	1
	Sn	IS	0.74	. . .	1
Steel	Cu	SPOF	. . .	0.80	5
	Pb	SPOF	. . .	1.40	5

Steel, 1032	Al, alloy 6061-T6	FOF	0.47	0.38	2
	Cu	FOF	0.32	0.25	2
	Steel, 1032	FOF	0.31	0.23	2
	Ti-6Al-4V	FOF	0.36	0.32	2
Steel, 52100	Ni_3Al, alloy IC-396M	RSOF	. . .	1.08	6
	Ni_3Al, alloy IC-50	RSOF	. . .	0.70	6
	Steel, 1015 annealed	BOR	. . .	0.74	7
	Steel, dual-phase DP-80	BOR	. . .	0.55	7
	Steel, O2 tool	BOR	. . .	0.49	7
Steel, mild	Steel, mild	BOR	. . .	0.62	3
Steel, M50 tool	Ni_3Al, alloy IC-50	RSOF	. . .	0.68	6
Steel, stainless	Steel, tool	POR	. . .	0.53	3
Steel, stainless 304	Cu	FOF	0.23	0.21	2
Stellite	Steel, tool	POR	. . .	0.60	3
Ti	Al	IS	0.54	. . .	1
	Steel, 17-4 stainless	POF	0.48	0.48	8
	Ti	POF	0.47	0.40	8
	Ti	FOF	0.55	. . .	1
	Ti-6Al-4V	POF	0.43	0.36	8
Ti-6Al-4V	Al, alloy 6061-T6	FOF	0.41	0.38	2
	Cu-Al (bronze)	POF	0.36	0.27	8
	Nitronic 60	POF	0.38	0.31	8
	Steel, 17-4 stainless	POF	0.36	0.31	8
	Steel, Type 440C stainless	POF	0.44	0.37	8
	Stellite 12	POF	0.35	0.29	8
	Stellite 6	POF	0.45	0.36	8
	Ta	POF	0.53	0.53	8
	Ti-6Al-4V	FOF	0.36	0.30	2
	Ti-6Al-4V	POF	0.36	0.31	8
W	Cu	IS	0.41	. . .	1
	Fe	IS	0.47	. . .	1
	W	IS	0.51	. . .	1
Zn	Cu	IS	0.56	. . .	1
	Fe	IS	0.55	. . .	1
	Zn	IS	0.75	. . .	1

(a) Test geometry codes: BOR, flat block pressed against the cylindrical surface of a rotating ring; FOF, flat surface sliding on another flat surface; IS, sliding down an inclined surface; POR, pin sliding against the cylindrical surface of a rotating ring; RSOF, reciprocating, spherically ended pin on a flat surface; SPOF, spherically ended pin on a flat coupon.

Table 2 Friction coefficient data for ceramics sliding on various materials

Specimens tested in air at room temperature

Material		Test geometry[(a)]	Friction coefficient		Ref
Fixed specimen	**Moving specimen**		**Static**	**Kinetic**	
Ag	Alumina	RPOF	. . .	0.37	9
	Zirconia	RPOF	. . .	0.39	9
Al	Alumina	RPOF	. . .	0.75	9
	Zirconia	RPOF	. . .	0.63	9
Alumina	Alumina	SPOD	. . .	0.50	10
	Alumina	SPOD	. . .	0.52	11
	Alumina	SPOD	. . .	0.33	12
	WRA[(b)]	SPOD	. . .	0.53	13
	WRZTA[(c)]	SPOD	. . .	0.50	13
	ZTA[(d)]	SPOD	. . .	0.56	13
Boron carbide	Boron carbide	POD	. . .	0.53	14
Cr	Alumina	RPOF	. . .	0.50	9
	Zirconia	RPOF	. . .	0.61	9
Cu	Alumina	RPOF	. . .	0.43	9

	Zirconia	RPOF	. . .	0.40	9
Fe	Alumina	RPOF		0.45	9
	Zirconia	RPOF	. . .	0.35	9
Glass, tempered	Al, alloy 6061-T6	FOF	0.17	0.14	15
	Steel, 1032	FOF	0.13	0.12	15
	Teflon[e]	FOF	0.10	0.10	15
Silicon carbide	Silicon carbide	SPOD	. . .	0.52	14
	Silicon nitride	SPOD	. . .	0.53	12
	Silicon nitride	SPOD	. . .	0.71	10
	Silicon nitride	SPOD	. . .	0.63	11
Silicon nitride	Silicon carbide	SPOD	. . .	0.54	12
	Silicon carbide	SPOD	. . .	0.67	10
	Silicon carbide	SPOD	. . .	0.84	11
	Silicon nitride	SPOD	. . .	0.17	14
Steel, M50 tool	Boron carbide	POD	. . .	0.29	14
	Silicon carbide	POD	. . .	0.29	14
	Silicon nitride	POD	. . .	0.15	14
	Tungsten carbide	POD	. . .	0.19	14
Ti	Alumina	RPOF	. . .	0.42	9
	Zirconia	RPOF	. . .	0.27	9
Tungsten carbide	Tungsten carbide	POD	. . .	0.34	14

(a) Test geometry codes: FOF, flat surface sliding on another flat surface; POD, pin on disk (pin tip geometry not given); RPOF, reciprocating pin on flat; SPOD, spherically ended pin on flat disk; SPOF, spherically ended pin on a flat coupon.

(b) WRA, silicon carbide whisker-reinforced alumina.

(c) WRZTA, silicon carbide whisker-reinforced, zirconia-toughened alumina.

(d) ZTA, zirconia-toughened alumina.

(e) Teflon, polytetrafluoroethylene

Table 3 Friction coefficient data for polymers sliding on various materials

Specimens tested in air at room temperature

Material[a]		Test geometry[b]	Friction coefficient		Ref
Fixed specimen	**Moving specimen**		**Static**	**Kinetic**	
Polymers sliding on polymers					
Acetal	Acetal	TW	0.06	0.07	16
Nylon 6/6	Nylon 6/6	TW	0.06	0.07	16
PMMA	PMMA	NSp	0.80	. . .	17
Polyester PBT	Polyester PBT	TW	0.17	0.24	16
Polystyrene	Polystyrene	NSp	0.50	. . .	17
Polyethylene	Polyethylene	NSp	0.20	. . .	17
Teflon	Teflon	NSp	0.04	. . .	17
	Teflon	FOF	0.08	0.07	18
Dissimilar pairs with the polymer as the fixed specimen					
Nylon 6 (cast)	Steel, mild	TPOD	. . .	0.35	19
(extruded)	Steel, mild	TPOD	. . .	0.37	19
Nylon 6/6	Polycarbonate	TW	0.25	0.04	16
Nylon 6/6 (+ PTFE)	Steel, mild	TPOD	. . .	0.35	19
PA 66	Steel, 52100	BOR	. . .	0.57	20
PA 66 (+ 15% PTFE)	Steel, 52100	BOR	. . .	0.13	20
PA 66 (PTFE/glass)	Steel, 52100	BOR	. . .	0.31	20
PEEK	Steel, 52100	BOR	. . .	0.49	20
PEEK (+ 15% PTFE)	Steel, 52100	BOR	. . .	0.18	20
PEEK (PTFE/glass)	Steel, 52100	BOR	. . .	0.20	20
PEI	Steel, 52100	BOR	. . .	0.43	20
PEI (+ 15% PTFE)	Steel, 52100	BOR	. . .	0.21	20
PEI (PTFE/glass)	Steel, 52100	BOR	. . .	0.21	20
PETP	Steel, 52100	BOR	. . .	0.68	20
PETP (+ 15% PTFE)	Steel, 52100	BOR	. . .	0.14	20

PETP (PTFE/glass)	Steel, 52100	BOR	. . .	0.18	20
Polyurethane[c]	Steel, mild	TPOD	. . .	0.51	19
Polyurethane[d]	Steel, mild	TPOD	. . .	0.35	19
POM	Steel, 52100	BOR	. . .	0.45	20
POM (+ 15% PTFE)	Steel, 52100	BOR	. . .	0.21	20
POM (PTFE/glass)	Steel, 52100	BOR	. . .	0.23	20
PPS	Steel, 52100	BOR	. . .	0.70	20
PPS (+ 15% PTFE)	Steel, 52100	BOR	. . .	0.30	20
PPS (PTFE/glass)	Steel, 52100	BOR	. . .	0.39	20
Teflon	Al, alloy 6061-T6	FOF	0.24	0.19	18
	Cr plate	FOF	0.09	0.08	18
	Cu	FOF	0.13	0.11	18
	Ni (0.001 P)	FOF	0.15	0.12	18
	Steel, 1032	FOF	0.27	0.27	18
	Ti-6Al-4V	FOF	0.17	0.14	18
	TiN (Magnagold)	FOF	0.15	0.12	18
UHMWPE	Steel, mild	TOPD	. . .	0.14	19
Dissimilar pairs with the polymer as the moving specimen					
Steel, carbon	ABS resin	POF	0.40	0.27	21
Steel, mild	ABS	TW	0.30	0.35	16
	ABS + 15% PTFE	TW	0.13	0.16	16
	Acetal	TW	0.14	0.21	16
Steel, 52100	Acetal	POD	. . .	0.31	22
	HDPE	POD	. . .	0.25	22
Steel, carbon	HDPE	POF	0.36	0.23	21
	LDPE	POF	0.48	0.28	21
Steel, 52100	Lexan 101	POD	. . .	0.60	22
Steel, mild	Nylon (amorphous)	TW	0.23	0.32	16
Steel, carbon	Nylon 6	POF	0.54	0.37	21
Steel, mild	Nylon 6	TW	0.22	0.26	16
Steel, carbon	Nylon 6/6	POF	0.53	0.38	21
Steel, mild	Nylon 6/6	TW	0.20	0.28	16
Steel, carbon	Nylon 6/10	POF	0.53	0.38	21
Steel, mild	Nylon 6/10	TW	0.23	0.31	16
	Nylon 6/12	TW	0.24	0.31	16
	PEEK (Victrex)	TW	0.20	0.25	16
Steel, carbon	Phenol formaldehyde	POF	0.51	0.44	21
Steel, 52100	PMMA	POD	. . .	0.68	22
Steel, carbon	PMMA	POF	0.64	0.50	21
Steel, mild	Polycarbonate	TW	0.31	0.38	16
	Polyesther PBT	TW	0.19	0.25	16
	Polyethylene	TW	0.09	0.13	16
Steel, carbon	Polyimide	POF	0.46	0.34	21
	Polyoxylmethylene	POF	0.30	0.17	21
	Polypropylene	POF	0.36	0.26	21
Steel, mild	Polypropylene	TW	0.08	0.11	16
Steel, carbon	Polystyrene	POF	0.43	0.37	21
Steel, mild	Polystyrene	TW	0.28	0.32	16
	Polysulfone	TW	0.29	0.37	16
Steel, carbon	PVC	POF	0.53	0.38	21
	PTFE	POF	0.37	0.09	21
Al, alloy 6061-T6	Teflon	FOF	0.19	0.18	18
Cr plate	Teflon	FOF	0.21	0.19	18
Glass, tempered	Teflon	FOF	0.10	0.10	18
Ni (0.001 P)	Teflon	FOF	0.22	0.19	18
Steel, 1032	Teflon	FOF	0.18	0.16	18
Ti-6Al-4V	Teflon	FOF	0.23	0.21	18
TiN (Magnagold)	Teflon	FOF	0.16	0.11	18

(a) ABS, acrylonitrile butadiene styrene; HDPE, high-density polyethylene; LPDE, low-density polyethylene; Lexan, trademark of the General Electric Co. (polycarbonate); nylon, one of a group of polyamide resins (see also PA); PA, polyamide; PBT, polybutylene terephthalate; PEEK, polyetheretherketone; PEI, polyetherimide; PETP, polyethylene terephthalate; PMMA, polymethylmethacrylate; POM, polyoxymethylene; PPS, polyphenylene

sulphide; PTFE, polytetrafluoroethylene; PVC, polyvinyl chloride; UHMWPE, ultra high molecular weight polyethylene; Magnagold, product of General Magnaplate, Inc.; Teflon, trademark of E.I. Du Pont de Nemours & Co., Inc. (PTFE).

(b) Test geometry codes: BOR, flat block-on-rotating ring; FOF, flat surface sliding on another flat surface; NSp, not specified; POD, pin on disk; POF, pin on flat; TPOD, triple pin-on-disk; TW, thrust washer test.

(c) Green polyurethane.

(d) Cream-colored polyurethane

Table 4 Friction coefficient data for coatings sliding on various materials

Specimens tested in air at room temperature

Material		Test geometry[a]	Friction coefficient		Ref
Fixed specimen	**Moving specimen**		**Static**	**Kinetic**	
Al, alloy 6061-T6	Cr plate	FOF	0.27	0.22	23
	Ni (0.001 P) plate	FOF	0.33	0.25	23
	TiN (Magnagold)[c]	FOF	0.25	0.22	23
Au, electroplate	60Pd-40Ag, plate	POF	. . .	2.40	24
	60Pd-40Au, plate	POF	. . .	0.30	24
	70Au-30Ag, plate	POF	. . .	3.00	24
	80Pd-20Au, plate	POF	. . .	1.80	24
	99Au-1 Co, plate	POF	. . .	2.40	24
	Au plate	POF	. . .	2.80	24
	Au-0.6 Co, plate	POF	. . .	0.40	24
	Pd plate	POF	. . .	0.60	24
Cr plate	Al, alloy 6061-T6	FOF	0.20	0.19	23
	Ni (0.001 P) plate	FOF	0.19	0.17	23
	Steel, 1032	FOF	0.20	0.17	23
	Teflon[b]	FOF	0.21	0.19	23
	Ti-6A1-4V	FOF	0.38	0.33	23
Niobium carbide, coating	Niobium carbide, coating	FOF	0.19	0.13	25
Ni (0.001 P) plate	Al, alloy 6061-T6	FOF	0.26	0.23	23
	Cr plate	FOF	0.41	0.36	23
	Ni (0.001 P) plate	FOF	0.32	0.28	23
	Steel, 1032	FOF	0.35	0.31	23
	Steel, D2 tool	FOF	0.43	0.33	23
	Teflon[b]	FOF	0.22	0.19	23
	TiN (Magnagold)[c]	FOF	0.33	0.26	23
Steel	Cu film on steel	SPOD	0.30	. . .	26
	In film on Ag	SPOD	0.10	. . .	26
	In film on steel	SPOD	0.08	. . .	26
	Pb film on Cu	SPOD	0.18	. . .	26
Steel, 1032	Cr plate	FOF	0.25	0.21	23
	Ni (0.001 P) plate	FOF	0.37	0.30	23
	TiN (Magnagold)[c]	FOF	0.31	0.28	23
Steel, type 440C stainless	TiC on type 304 stainless	POD	0.12	0.17	27
	TiN on type 304 stainless	POD	0.50	0.75	27
Steel, bearing	Chrome carbide	POD	. . .	0.79	28
	SiC (CVD)[d]	POD	. . .	0.23	28
	TiC (CVD)[d]	POD	. . .	0.25	28
	TiN (CVD)[d]	POD	. . .	0.49	28
Steel, stainless	Al_2O_3, plasma-sprayed	Ams	. . .	0.13-0.30	29
	Cr plate	Ams	. . .	0.30-0.38	29
	Cr_2O_3, plasma-sprayed	Ams	. . .	0.14-0.15	29
	TiO_2, plasma sprayed	Ams	. . .	0.10-0.15	29
	WC-12 Co, plasma-sprayed	Ams	. . .	0.11-0.13	29
Teflon[b]	Cr plate	FOF	0.09	0.08	23
	Ni (0.001 P) plate	FOF	0.15	0.12	23
	TiN (Magnagold)[c]	FOF	0.15	0.12	23
TiC on type 44OC stainless steel	Al	POD	0.50	0.85	27

	Ti	POD	0.65	0.80	27
	TiC on type 440C stainless steel	POD	0.22	0.20	27
	TiN on type 440C stainless steel	POD	0.25	0.20	27
TiN on type 440C stainless steel	Al	POD	0.27	0.40	27
	Steel, type 304 stainless	POD	0.29	0.41	27
	Ti	POD	0.50	0.76	27
	TiC on type 440C stainless steel	POD	0.05	0.06	27
	TiN on type 440C stainless steel	POD	0.65	0.45	27
TiN (Magnagold)[(c)]	Al, alloy 6061-T6	FOF	0.30	0.26	23
	Steel, 1032	FOF	0.38	0.31	23
	Teflon[(b)]	FOF	0.16	0.11	23
	Ti-6Al-4V	FOF	0.26	0.23	23
	TiN (Magnagold)[(c)]	FOF	0.25	0.21	23

(a) Ams, Amsler circumferential, rotating disk-on-disk machine; FOF, flat surface sliding on another flat surface; POD, pin on disk; POF, pin on flat; SPOD, spherically ended pin-on-flat disk.

(b) Teflon is a registered trademark of E.I. Du Pont de Nemours & Co., Inc. (polytetrafluoroethylene).

(c) Magnagold is a product of General Magnaplate, Inc.

(d) CVD, chemical vapor deposition

Table 5 Friction coefficient data for miscellaneous materials

Specimens tested in air at room temperature

Material		Test geometry[(a)]	Friction coefficient		Ref
Fixed specimen	**Moving specimen**		**Static**	**Kinetic**	
Brick	Wood	UnSp	0.6	. . .	30
Cotton thread	Cotton thread	UnSp	0.3	. . .	30
Diamond	Diamond	UnSp	0.1	. . .	30
Explosives[(b)]					
HMX[(c)]	Glass	RPOF	. . .	0.55	31
PETN[(d)]	Glass	RPOF	. . .	0.40	31
RDX[(e)]	Glass	RPOF	. . .	0.35	31
Lead azide [$Pb(N_3)_2$]	Glass	RPOF	. . .	0.28	31
Silver azide (AgN_3)	Glass	RPOF	. . .	0.40	31
Glass, tempered	Al, alloy 6061-T6	FOF	0.17	0.14	32
	Steel, 1032	FOF	0.13	0.12	32
	Teflon[(f)]	FOF	0.10	0.10	32
Glass, thin fiber	Brass	StOD	. . .	0.16-0.26	33
	Graphite	StOD	. . .	0.15	33
	Porcelain	StOD	. . .	0.36	33
	Steel, stainless	StOD	. . .	0.31	33
	Teflon[(f)]	StOD	. . .	0.10	33
Glass, clean	Glass (clean)	UnSp	0.9-1.0	. . .	30
Graphite, molded	Al, alloy 2024	FOF	0.16	. . .	34
	Al, alloy 2219	FOF	0.22	. . .	34
	Graphite, extruded	FOF	0.20	0.17	34
	Graphite, molded	FOF	0.18	0.14	34
	Inconel X-750[(g)]	FOF	0.16	. . .	34
	Steel, type 304 stainless	FOF	0.18	. . .	34
	Steel, type 347 stainless	FOF	0.19	. . .	34
Graphite (clean)	Graphite (clean)	UnSp	0.10	. . .	30
Graphite (outgassed)	Graphite (outgassed)	UnSp	0.5-0.8	. . .	30
Hickory wood, waxed	Snow	UnSp	. . .	0.14	35
Ice	Bronze	UnSp	. . .	0.02	35
	Ebonite	UnSp	. . .	0.02	35
	Ice	UnSp	0.05-0.15	. . .	35
	Ice	UnSp	. . .	0.02	35
	Ice	FOF	>0.01	>0.01	32
Leather	Metal (clean)	UnSp	0.6	. . .	30
Metal	Glass (clean)	UnSp	0.5-0.7	. . .	30

Mica (cleaved)	Mica (cleaved)	UnSp	1.0	. . .	30
Mica (contaminated)	Mica (contaminated)	UnSp	0.2-0.4	. . .	30
Nylon fibers	Nylon fibers	UnSp	0.15-0.25	. . .	30
Paper, copier	Paper, copier	FOF	0.28	0.26	32
Sapphire	Sapphire	UnSp	0.2	. . .	30
Silk fibers	Silk fibers	UnSp	0.2-0.3	. . .	30
Steel (clean)	Graphite	UnSp	0.1	. . .	30
Wood (clean)	Metals	UnSp	0.2-0.6	. . .	30
	Wood (clean)	UnSp	0.25-0.5	. . .	30

(a) FOF, flat surface sliding on another flat surface; RPOF, reciprocating pin-on-flat; StOD, strand wrapped over a drum; UnSp, unspecified method.

(b) Explosives reported here were tested as reciprocating, single-crystal, flat-ended pin-on-moving flat.

(c) HMX, cyclotetramethylene tetranitramine;

(d) PETN, pentaerithritol tetranitrate.

(e) RDX, cyclotrimethylene trinitramine.

(f) Teflon is a registered trademark of E.I. Du Pont de Nemours & Co., Inc.

(g) Inconel is a product of INCO, Inc.

References

1. E. Rabinowicz, *ASLE Trans.*, Vol 14, 1971, p 198; plate sliding on plate at 50% relative humidity
2. "Friction Data Guide," General Magnaplate Corporation, 1988; TMI Model 98-5 slip and friction tester, 1.96 N (0.200 kgf) load, ground specimens, 54% relative humidity, average of five tests
3. J.F. Archard, *ASME Wear Control Handbook*, M.B. Peterson and W.O. Winer, Ed., American Society of Mechanical Engineers, 1980, p 38; pin-on-rotating ring, 3.9 N (0.40 kgf) load, 1.8 m/s (350 ft/min) velocity
4. A.W. Ruff, L.K. Ives, and W.A. Glaeser, *Fundamentals of Friction and Wear of Materials,* ASM International, 1981, p 235; flat block-on-rotating 35 mm ($1\frac{3}{8}$ in.) diameter ring, 10 N (1.02 kgf) load, 0.2 m/s (40 ft/min) velocity
5. F.P. Bowden and D. Tabor, *The Friction and Lubrication of Solids,* Oxford Press, 1986, p 127; sphere-on-flat, unspecified load and velocity
6. P.J. Blau and C.E. DeVore, *Tribol. Int.,* Vol 23 (No. 4), 1990, p 226; reciprocating ball-on-flat, 10 Hz, 25 N (2.6 kgf) load, 10 mm stroke
7. P.J. Blau, *J. Tribology,* Vol 107, 1985, p 483; flat block-on-rotating 35 mm ($1\frac{3}{8}$ in.) diameter ring, 133 N (13.6 kgf) load, 5.0 cm/s (2.0 in./s) velocity
8. K.G. Budinski, *Proceedings of Wear of Materials,* American Society of Mechanical Engineers, 1991, p 289; modified ASTM G 98 galling test procedure
9. K. Demizu, R. Wadabayashim, and H. Ishigaki, *Tribol. Trans.*, Vol 33 (No. 4), 1990, p 505; 1.5 mm (0.060 in.) radius pin reciprocating on a flat, 4 N (0.4 kgf) load, 0.17 mm/s (0.0067 in./s) velocity, 50% relative humidity
10. P.J. Blau, Oak Ridge National Laboratory
11. P.J. Blau, Oak Ridge National Laboratory, 1.0 N (0.10 kgf) load and 0.1 m/s (20 ft/min) velocity
12. P.J. Blau, Oak Ridge National Laboratory, 10 N (1.0 kgf) load and 0.1 m/s (20 ft/min) velocity
13. C.S. Yust, *Tribology of Composite Materials,* P.K. Rohatgi, P.J. Blau, and C.S. Yust, Ed., ASM International, 1990, p 27; 9.5 mm ($\frac{3}{8}$ in.) diameter sphere-on-disk, 2 to 9 N (0.2 to 0.9 kgf) load, 0.3 m/s (60 ft/min) velocity
14. B. Bhushan and B.K. Gupta, table in *Handbook of Tribology,* Mc-Graw-Hill, 1991; 20 N (2.0 kgf), 3 mm/s (0.12 in./s) velocity
15. "Friction Data Guide," General Magnaplate Corporation, 1988; TMI Model 98-5 slip and friction tester, 1.96 N (0.200 kgf) load, ground specimens, 54% relative humidity, average of five tests

16. "Lubricomp(R) Internally-Lubricated Reinforced Thermoplastics and Fluoropolymer Composites," Bulletin 254-688, ICI Advanced Materials; thrust washer apparatus, 0.28 MPa (40 psi), 0.25 m/s (50 ft/min), after running-in for one full rotation
17. F.P. Bowden and D. Tabor, Appendix IV, *The Friction and Lubrication of Solids,* Oxford Press, 1986; unspecified testing conditions
18. "Friction Data Guide," General Magnaplate Corporation, 1988; TMI Model 98-5 slip and friction tester, 1.96 N (0.200 kgf) load, ground specimens, 54% relative humidity, average of five tests
19. J.M. Thorp, *Tribol. Int.,* Vol 15 (No. 2), 1982, p 69; three-pin-on-rotating disk apparatus, 0.1 m/s (20 ft/min)
20. J.W.M. Mens and A.W.J. de Gee, *Wear,* Vol 149, 1991, p 255; flat block-on-rotating ring, 1.5 MPa (0.22 ksi) pressure, 150 N (15 kgf) load, 0.1 m/s (20 ft/min) velocity
21. R.P. Steijn, *Metall. Eng. Quart.,* Vol 7, 1967, p 9; 12.7 mm (0.500 in.) diameter ball-on-flat, 9.8 N (1.0 kgf) load, 0.01 mm/s (4×10^{-4} in./s) velocity
22. N.P. Suh, *Tribophysics,* Prentice-Hall, 1986, p 226; pin-on-disk, 4.4 N (0.45 kgf) load, 3.3 cm/s (1.3 in./s) velocity, 65% relative humidity
23. "Friction Data Guide," General Magnaplate Corporation, 1988; TMI Model 98-5 slip and friction tester, 1.96 N (0.200 kgf) load, ground specimens, 54% relative humidity, average of five tests
24. M. Antler and E.T. Ratcliff, *Proceedings of the Holm Conference on Electrical Contacts,* 1982, p 19; sphere-on-reciprocating flat, 0.49 N (0.050 kgf) load, 1.0 mm/s (0.039 in./s) velocity
25. M.J. Manjoine, *Bearing and Seal Design in Nuclear Power Machinery,* American Society of Mechanical Engineers, 1967; flat plate-on-flat plate, 28 MPa (4.1 ksi) contact pressure, 0.25 mm/s (0.010 in./s) velocity
26. F.P. Bowden and D. Tabor, *The Friction and Lubrication of Solids,* Oxford Press, 1986, p 127; sphere-on-flat, low-speed sliding, 39.2 N (4 kgf) load
27. B. Bhushan and B.K. Gupta, *Handbook of Tribology,* McGraw-Hill, 1991, Table 14.16a; pin-on-disk, 12 N (1.2 kgf) load, 14 to 16 cm/s (0.55 to 0.63 in./s) velocity
28. B. Bhushan and B.K. Gupta, *Handbook of Tribology,* McGraw-Hill, 1991, Table 14.65; pin-on-disk, 5 N (0.5 kgf) load, 1.0 cm/s (0.39 in./s) velocity, 50% relative humidity
29. B. Bhushan and B.K. Gupta, *Handbook of Tribology,* McGraw-Hill, 1991, Table 14.12; Amsler disk machine, 400 rev/min, 250 N (26 kgf) load
30. F.P. Bowden and D. Tabor, Appendix IV, *The Friction and Lubrication of Solids,* Oxford Press, 1986; method unspecified
31. J.K.A. Amuzu, B.J. Briscoe, and M.M. Chaudhri, *J. Phys. D, Appl. Phys.,* Vol 9, 1976, p 133; reciprocating, single-crystal flat sliding on smooth fired glass surfaces, range 5 to 20 gf (0.049 to 0.1962 N load), 0.20 mm/s (0.008 in./s) velocity
32. "Friction Data Guide" General Magnaplate Corporation, 1988; TMI Model 98-5 slip and friction tester, 1.96 N (0.200 kgf) load, ground specimens, 54% relative humidity, average of five tests
33. P.K. Gupta, *J. Am. Ceram. Soc.,* Vol 74 (No. 7), 1991, p 1692; strand lying on a rotating drum, 1.96 N (0.200 kgf) load, 8.5 mm/s (0.33 in./s) velocity
34. M.J. Manjoine, *Bearing and Seal Design in Nuclear Power Machinery,* American Society of Mechanical Engineers, 1967; flat plate-on-flat plate, 28 MPa (4.1 ksi) contact pressure, 0.25 mm/s (0.010 in./s) velocity
35. F.P Bowden and D. Tabor, *The Friction and Lubrication of Solids,* Oxford Press, 1986; unspecified method, 4.0 m/s (790 ft/min) at 0 °C

References

REFERENCES/TEST CONDITIONS (Table 1 Friction coefficient data for metals sliding on metals)

1. E. Rabinowicz, *ASLE Trans.,* Vol 14, 1971, p 198; plate sliding on plate at 50% relative humidity

2. "Friction Data Guide," General Magnaplate Corporation, 1988; TMI Model 98-5 slip and friction tester, 1.96 N (0.200 kgf) load, ground specimens, 54% relative humidity, average of five tests
3. J.F. Archard, *ASME Wear Control Handbook,* M.B. Peterson and W.O. Winer, Ed., American Society of Mechanical Engineers, 1980, p 38; pin-on-rotating ring, 3.9 N (0.40 kgf) load, 1.8 m/s (350 ft/min) velocity
4. A.W. Ruff, L.K. Ives, and W.A. Glaeser, *Fundamentals of Friction and Wear of Materials,* ASM International, 1981, p 235; flat block-on-rotating 35 mm (1 $\frac{3}{8}$ in.) diameter ring, 10 N (1.02 kgf) load, 0.2 m/s (40 ft/min) velocity
5. F.P. Bowden and D. Tabor, *The Friction and Lubrication of Solids,* Oxford Press, 1986, p 127; sphere-on-flat, unspecified load and velocity
6. P.J. Blau and C.E. DeVore, *Tribol. Int.,* Vol 23 (No. 4), 1990, p 226; reciprocating ball-on-flat, 10 Hz, 25 N (2.6 kgf) load, 10 mm stroke
7. P.J. Blau, *J. Tribology,* Vol 107, 1985, p 483; flat block-on-rotating 35 mm (1 $\frac{3}{8}$ in.) diameter ring, 133 N (13.6 kgf) load, 5.0 cm/s (2.0 in./s) velocity
8. K.G. Budinski, *Proceedings of Wear of Materials,* American Society of Mechanical Engineers, 1991, p 289; modified ASTM G 98 galling test procedure

REFERENCES/TEST CONDITIONS (Table 2 Friction coefficient data for ceramics sliding on various materials)

9. K. Demizu, R. Wadabayashim, and H. Ishigaki, *Tribol. Trans.,* Vol 33 (No. 4), 1990, p 505; 1.5 mm (0.060 in.) radius pin reciprocating on a flat, 4 N (0.4 kgf) load, 0.17 mm/s (0.0067 in./s) velocity, 50% relative humidity
10. P.J. Blau, Oak Ridge National Laboratory
11. P.J. Blau, Oak Ridge National Laboratory, 1.0 N (0.10 kgf) load and 0.1 m/s (20 ft/min) velocity
12. P.J. Blau, Oak Ridge National Laboratory, 10 N (1.0 kgf) load and 0.1 m/s (20 ft/min) velocity
13. C.S. Yust, *Tribology of Composite Materials,* P.K. Rohatgi, P.J. Blau, and C.S. Yust, Ed., ASM International, 1990, p 27; 9.5 mm ($\frac{3}{8}$ in.) diameter sphere-on-disk, 2 to 9 N (0.2 to 0.9 kgf) load, 0.3 m/s (60 ft/min) velocity
14. B. Bhushan and B.K. Gupta, table in *Handbook of Tribology,* Mc-Graw-Hill, 1991; 20 N (2.0 kgf), 3 mm/s (0.12 in./s) velocity
15. "Friction Data Guide," General Magnaplate Corporation, 1988; TMI Model 98-5 slip and friction tester, 1.96 N (0.200 kgf) load, ground specimens, 54% relative humidity, average of five tests

REFERENCES/TEST CONDITIONS (Table 3 Friction coefficient data for polymers sliding on various materials)

16. "Lubricomp(R) Internally-Lubricated Reinforced Thermoplastics and Fluoropolymer Composites," Bulletin 254-688, ICI Advanced Materials; thrust washer apparatus, 0.28 MPa (40 psi), 0.25 m/s (50 ft/min), after running-in for one full rotation
17. F.P. Bowden and D. Tabor, Appendix IV, *The Friction and Lubrication of Solids,* Oxford Press, 1986; unspecified testing conditions
18. "Friction Data Guide," General Magnaplate Corporation, 1988; TMI Model 98-5 slip and friction tester, 1.96 N (0.200 kgf) load, ground specimens, 54% relative humidity, average of five tests
19. J.M. Thorp, *Tribol. Int.,* Vol 15 (No. 2), 1982, p 69; three-pin-on-rotating disk apparatus, 0.1 m/s (20 ft/min)
20. J.W.M. Mens and A.W.J. de Gee, *Wear,* Vol 149, 1991, p 255; flat block-on-rotating ring, 1.5 MPa (0.22 ksi) pressure, 150 N (15 kgf) load, 0.1 m/s (20 ft/min) velocity
21. R.P. Steijn, *Metall. Eng. Quart.,* Vol 7, 1967, p 9; 12.7 mm (0.500 in.) diameter ball-on-flat, 9.8 N (1.0 kgf) load, 0.01 mm/s (4×10^{-4} in./s) velocity
22. N.P. Suh, *Tribophysics,* Prentice-Hall, 1986, p 226; pin-on-disk, 4.4 N (0.45 kgf) load, 3.3 cm/s (1.3 in./s) velocity, 65% relative humidity

REFERENCES/TEST CONDITIONS (Table 4 Friction coefficient data for coatings sliding on various materials)

23. "Friction Data Guide," General Magnaplate Corporation, 1988; TMI Model 98-5 slip and friction tester, 1.96 N (0.200 kgf) load, ground specimens, 54% relative humidity, average of five tests
24. M. Antler and E.T. Ratcliff, *Proceedings of the Holm Conference on Electrical Contacts,* 1982, p 19; sphere-on-reciprocating flat, 0.49 N (0.050 kgf) load, 1.0 mm/s (0.039 in./s) velocity
25. M.J. Manjoine, *Bearing and Seal Design in Nuclear Power Machinery,* American Society of Mechanical Engineers, 1967; flat plate-on-flat plate, 28 MPa (4.1 ksi) contact pressure, 0.25 mm/s (0.010 in./s) velocity
26. F.P. Bowden and D. Tabor, *The Friction and Lubrication of Solids,* Oxford Press, 1986, p 127; sphere-on-flat, low-speed sliding, 39.2 N (4 kgf) load
27. B. Bhushan and B.K. Gupta, *Handbook of Tribology,* McGraw-Hill, 1991, Table 14.16a; pin-on-disk, 12 N (1.2 kgf) load, 14 to 16 cm/s (0.55 to 0.63 in./s) velocity
28. B. Bhushan and B.K. Gupta, *Handbook of Tribology,* McGraw-Hill, 1991, Table 14.65; pin-on-disk, 5 N (0.5 kgf) load, 1.0 cm/s (0.39 in./s) velocity, 50% relative humidity
29. B. Bhushan and B.K. Gupta, *Handbook of Tribology,* McGraw-Hill, 1991, Table 14.12; Amsler disk machine, 400 rev/min, 250 N (26 kgf) load

REFERENCES/TEST CONDITIONS (Table 5 Friction coefficient data for miscellaneous materials)

30. F.P. Bowden and D. Tabor, Appendix IV, *The Friction and Lubrication of Solids,* Oxford Press, 1986; method unspecified
31. J.K.A. Amuzu, B.J. Briscoe, and M.M. Chaudhri, *J. Phys. D, Appl. Phys.*, Vol 9, 1976, p 133; reciprocating, single-crystal flat sliding on smooth fired glass surfaces, range 5 to 20 gf (0.049 to 0.1962 N load), 0.20 mm/s (0.008 in./s) velocity
32. "Friction Data Guide" General Magnaplate Corporation, 1988; TMI Model 98-5 slip and friction tester, 1.96 N (0.200 kgf) load, ground specimens, 54% relative humidity, average of five tests
33. P.K. Gupta, *J. Am. Ceram. Soc.,* Vol 74 (No. 7), 1991, p 1692; strand lying on a rotating drum, 1.96 N (0.200 kgf) load, 8.5 mm/s (0.33 in./s) velocity
34. M.J. Manjoine, *Bearing and Seal Design in Nuclear Power Machinery,* American Society of Mechanical Engineers, 1967; flat plate-on-flat plate, 28 MPa (4.1 ksi) contact pressure, 0.25 mm/s (0.010 in./s) velocity
35. F.P Bowden and D. Tabor, *The Friction and Lubrication of Solids,* Oxford Press, 1986; unspecified method, 4.0 m/s (790 ft/min) at 0 °C

Introduction to Lubrication

Herbert S. Cheng, Northwestern University

LUBRICATION is introduced between two sliding solids by adding a gaseous, liquid, or solid lubricant at the sliding interface in order to reduce friction and wear, and to carry away heat and debris generated during the sliding process. Lubrication processes can take many different forms, depending on the gross geometry of the contacting bodies, the roughness and texture of the sliding surfaces, the contacting load, the pressure and temperature, the rolling and sliding speeds, the environmental conditions, the physical and chemical properties of the lubricant, the material composition, and the properties of the near-surface layer.

The two basic geometries for lubricated surfaces are conformal and counterformal (Fig. 1). Conformal surfaces usually appear in sliding journal and trust bearings, machine guideways, and seals. In conformal bearings, surfaces are usually separated by a thick oil or gas film generated hydrodynamically by the surface velocities or hydrostatically by an externally pressurized lubricant. These surfaces usually operate in the regime of thick-film hydrodynamic or hydrostatic lubrication. Lubrication performance--as measured by film thickness, pressure, and friction--can be predicted readily from analyses based on the Reynolds equation (Ref 1). The major lubricant properties of importance for conformal surfaces include viscosity and temperature-viscosity coefficient. Typical conformal bearings include engine main-shaft bearings; turbine; compressor, and pump bearings; and computer head/disk contacts.

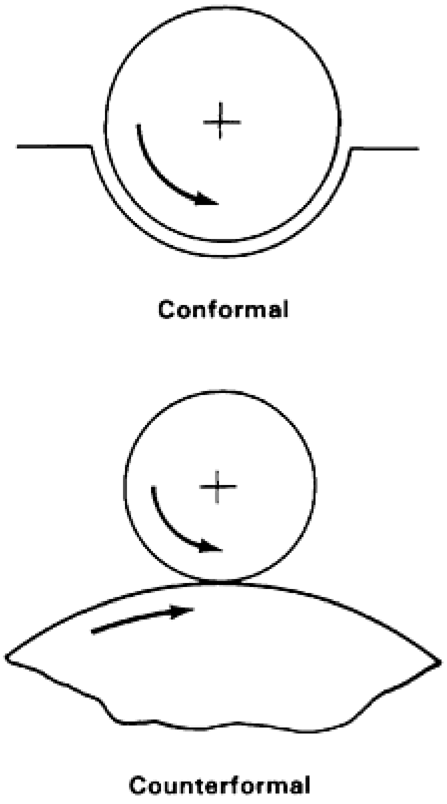


Fig. 1 Geometry of conformal and counterformal contacts

Lubrication of counterformal contacts takes place in an extremely small concentrated area known as the Hertzian conjunction. In these contacts, film thickness is very thin, of the same order of the surface roughness, and lubricant pressure is very high. Lubricant performance is affected strongly by the elastic deformation of the bearing surface. The lubricant film thickness and pressure and their distribution in the conjunction can be determined by elastohydrodynamic theories (Ref 2). Because of the extremely high pressure in the conjunction, the pressure-viscosity coefficient has a major

effect in generating the lubricant film at the inlet, and the temperature-viscosity coefficient has a strong influence on both temperature and friction for sliding contacts. These significant lubrication properties are discussed in more detail in the article "Liquid Lubricants" in this Section. Recent molecular-level studies indicate that the structure and properties of the first few molecular layers of a lubricant can be different from those of the bulk liquid.

Surface roughness and its lay orientation with respect to surface motion can have a significant influence on lubrication performance in the regime of mixed lubrication, where the load is shared between the lubricant pressure and the asperities. Roughness effects are particularly important in counterformal contacts because the majority of these contacts operate in mixed lubrication. Progressive wear failure and sudden scuffing failure are dependent on the lubrication process influenced by the micro-roughness geometry; they are also influenced by the contact temperatures in the conjunction. Detailed information about mixed lubrication can be found in the article "Lubrication Regimes" in this Section; contact temperature calculations are discussed in the article "Frictional Heating Calculations" in this Volume.

Thick-film lubrication deteriorates rapidly under combinations of extreme pressure, extreme temperature, low rolling speed, and high sliding speed. For these conditions, the lubricant film thickness is extremely small in comparison to the surface roughness, and the contacts operate in the regime of boundary lubrication. Straight mineral oils do not offer much relief from friction and wear. Additives must be relied upon to modify friction, reduce wear, and reduce oxidation of the lubricant. The functions of commonly used additives are explained in the article "Lubricant Additives and Their Functions" in this Section.

In extreme environmental conditions, such as elevated ambient temperatures above 500 °C (930 °F) or vacuum environments, conventional liquid lubricants often become less effective because of their tendency to rapidly oxidize or decompose at elevated temperatures and their tendency to vaporize or creep away from lubricated surfaces under high vacuum. The use of solid lubricants is often called for in these situations. Over the last four decades, the use of solid lubrication has grown steadily, largely due to sustaining efforts at the NASA Lewis Research Center and research supported by the U.S. Air Force. These efforts are discussed in the articles "Solid Lubricants" and "Lubricants for High-Vacuum Applications" in this Section.

Lubrication in counterformal contacts also is influenced by the ratio of the sliding velocity to the rolling velocity, otherwise known as the slide-to-roll ratio. The film thickness is generated primarily by the rolling velocity, which draws the lubricant into the conjunction. Sliding velocity has no effect on film generation unless the inlet viscosity is reduced significantly by the heat generated by sliding. The friction and contact temperature in the conjunction is influenced strongly by the sliding velocity. A lubricant in sliding contact usually behaves like a sheet of glassy solid sheared at the midlayer. Contacts that operate at a high slide-to-roll ratio generate a higher contact surface temperature in the conjunction and have a lower resistance to scuffing. This condition exists in gear teeth contacts at the beginning of the engagement between a pair of teeth; in cam follower/lobe contacts at the beginning and end of the lifting process; and in piston ring/cylinder bore contacts throughout the entire reciprocation cycle. For these conditions, additives should be used to reduce asperity friction and thereby prevent the excessive asperity temperature rise that leads to scuffing.

Contacts that operate at very low slide-to-roll ratios (that is, nearly pure rolling condition) generate much less heat in the conjunction, and can support a higher load without surface distress generated by sliding wear. This condition exists in gear teeth contacts at or near the pitch point and in all contacts between raceway/ball or raceway/roller contacts of rolling-element bearings. Surface or subsurface fatigue is the predominant mode of failure. Surface fatigue life improves greatly with a higher ratio of lubricant film to surface roughness. Elastohydrodynamic lubrication theories are directly useful for assessing fatigue life of these contacts. Film thickness formulas can be found in the article "Lubrication Regimens" in this Section and in the article "Friction, Lubrication, and Wear of Gears," which appears later in this Volume. The function and selection of lubricants for meeting the requirements of specific triboelements are discussed in the articles "Lubricants for Rolling Element Bearings" and "Internal Combustion Engine Lubricants" in this Section.

In metalworking processes, lubrication takes place at the interface between an elastic solid and a plastically deformed solid. Thick-film lubrication can be achieved but often is not because of the poor surface quality generated under these conditions. Mixed or boundary lubrication is usually specified for most metalworking processes in order to achieve the desirable friction that will enable control of the surface quality of the workpiece. In addition, the lubricant must also be a good coolant to carry away the heat generated by plastic deformation of the workpiece. Detailed coverage of the formulation and properties of modern metalworking lubricants can be found in the article "Metalworking Lubricants" in this Section.

References

1. A.Z. Szeri, *Tribology: Friction, Lubrication, and Wear,* McGraw-Hill, 1980
2. G. Ramsey, *Elastohydrodynamics,* John Wiley & Sons, 1988

Liquid Lubricants

Richard S. Fein, Fein Associates

Introduction

THE PRIMARY FUNCTION of liquid lubricants is to control friction, wear, and surface damage over the intended life of a system that contains machine elements, such as gears and bearings. Wear and surface damage occur under boundary or partial boundary lubrication conditions, but not under full hydrodynamic conditions. Secondary functions are to prevent corrosion and to scavenge heat, dirt, and wear debris. Lubricants can also transfer either force or energy, as occurs in hydraulic systems.

The choice of an appropriate lubricant depends on system needs and cost, as well as on health, safety, and environmental considerations. For a given system, minimizing the number of lubricants used generally simplifies maintenance and reduces the chances of applying the wrong lubricant. Inventory storage and handling are also reduced.

Machinery manufacturers help define system needs by recommending viscosity grades and the lubricant quality that is suitable for use with their equipment in various applications. In some cases, rather detailed specifications governing the oils to be used may be provided. With new equipment, the recommendations of manufacturers should be followed to ensure warranty protection. After the warranty period, the number of different lubricants can often be simplified (Ref 1), with the assistance of an experienced lubricant supplier.

This article describes types of base oils and the properties important to lubricant functioning; lubricant classifications; health, safety, and environmental considerations; and methods of lubricant application. Hydrocarbon-based oils, especially mineral oils, are emphasized because they are the most widely used. Additional details on most of these topics are provided in Ref 1, 2, 3, 4, 5.

Properties

Lubricant properties depend on the base oil (or blend of base oils) and on any additives that are used. Table 1 qualitatively indicates the important characteristics of a variety of base oils. Overall, mineral oils have good characteristics and are low in cost.

Table 1 Base oil characteristics

Property	Base oil type									
	Mineral oil	**Olefin polymer**	**Alkyl aromatic**	**Polyphenyl ether**	**Dibasic acid ester**	**Neopentyl polyester**	**Poly-alkylene glycol**	**Phosphate ester**	**Silicone**	**Silicate ester**	**Fluoro-carbon**
Liquid range	Moderate	Good	Good	Poor	Very good	Very good	Good	Moderate	Excellent	Poor	Poor
Viscosity-temperature	Moderate	Good	Moderate	Poor	Excellent	Very good	Good	Poor	Excellent	Excellent	Moderate
Low-temperature flow	Poor	Good	Good	Poor	Good	Good	Good	Moderate	Good	Moderate	Good
Oxidation stability inhibited	Moderate	Very good	Moderate	Very good	Very good	Moderate	Poor	Good	Very good	Very good	Excellent
Hydrolytic stability	Excellent	Excellent	Excellent	Excellent	Moderate	Moderate	Good	Moderate	Good	Poor	Very good
Thermal stability	Moderate	Moderate	Moderate	Excellent	Good	Good	Good	Moderate	Very good	Good	Very good
Solvency, mineral oil	. . .	Excellent	Excellent	Good	Good	Moderate	Poor	Moderate	Poor	Moderate	Poor
Solvency, additives	Excellent	Good	Excellent	. . .	Very good	Very good	Moderate	Good	Poor	. . .	. . .
Solvency, varnish and paint	Excellent	Excellent	Excellent	Moderate	Good	Moderate	Moderate	Poor	Good	Moderate	Good
Volatility	Moderate	Good	Good	Good	Excellent	Excellent	Good	Good	Good	Good	Moderate
Antirust, inhibited	Excellent	Excellent	Excellent	. . .	Moderate	Moderate	Good	Moderate	Good	. . .	. . .
Boundary lubrication	Good	Good	Good	Excellent	Very Good	Very Good	Good	Excellent	Moderate	Moderate	Excellent
Fire resistance	Poor	Poor	Poor	Moderate	Moderate	Moderate	Moderate	Excellent	Moderate	Moderate	Excellent
Elastomer swell, Buna	Low	None	Low	Low	Medium	High	Low	High	Low	Low	Medium
Cost	$	$$	$$	$$$$	$$	$$	$$	$$	$$$	$$$	$$$$

Additives are chemicals that enhance the properties of an oil. They are added to base oils in concentrations that range from the order of tens of ppm to as much as 25%. Additives generally fall into three groups: polymers, polar compounds, and compounds containing active elements, such as sulfur and chlorine. They can also be classified by function, and are described as such in the article "Lubricant Additives and Their Functions" in this Volume. Some additives that offer advantages in one performance area may be detrimental in other areas, whereas others work best in combination. Additive use is justified by cost effectiveness.

The numerous lubricant properties that are important to end-use applications are listed in one of two categories in Table 2. Basic properties are physical or chemical characteristics that are essentially independent of the measurement method. Performance properties, on the other hand, do depend on the measurement method, because they are governed by interacting basic properties and test conditions. The most important properties, as well as those most often used for lubricant characterization, are described below, and typical values for selected properties are given in Table 3.

Table 2 Lubricant properties

Basic properties
Bulk modulus
Surface tension
Thermal conductivity
Electrical conductivity
Heat capacity
Refractive index
Rheology
Viscosity
Viscosity-pressure coefficient
Viscosity-temperature coefficient
Temporary viscosity loss
Gravity
Density
Performance properties
Corrosion/rust
Wettability
Volatility
Foaming/air release
Toxicity
Biodegradability
Miscibility
Emulsibility
Filterability
Compatibility
Seals
Paint
Flammability
Rheology
Shear strength
Pourability
Permanent viscosity loss
Tribology
Friction
Wear
Load carrying
Pitting fatigue
Solvency
Stability
Oxidative
Hydrolytic
Thermal
Detergency/dispersancy

Table 3 Lubricant properties

Lubricant	Pour point °C	Pour point °F	API gravity	Flash point °C	Flash point °F	Kinematic viscosity, mm^2/s, cSt At 40 °C (105 °F)	Kinematic viscosity, mm^2/s, cSt At 100 °C (212 °F)	Viscosity index	Roelands, Z	Ash, %	Neutrality No.
Jet engine oil, polyol ester, type II, MIL-L-23699B	-54	-65	10.0	254	489	28.0	5.10	. . .	0.61	. . .	. . .
ISO VG 32 refrigeration oil	-48	-55	27.2	168	334	30.0	4.70	54	0.76	. . .	0.03 TAN
ISO VG 32 R&O turbine oil	-32	-26	32.2	199	390	30.9	5.48	104	0.68	0.002	. . .
Automatic transmission fluid, Dexron-II, Mercon, Allison C-3, Caterpillar TO-2	-46	-51	31.0	204	399	36.2	7.40	176	0.54	0.26	. . .
ISO VG 46 antiwear hydraulic fluid, AGMA No. 1	-32	-26	30.2	210	410	45.0	6.80	103	0.67	. . .	. . .
5W-30 motor oil, SG/CD, energy conserving II, ILSAC, CCMC	-39	-38	30.0	216	421	67.8	11.5	165	0.51	1.0	. . .
10W-30 motor oil, SG/CD, ILSAC, energy conserving II, CCMC	-34	-29	29.0	221	430	75.0	11.5	146	0.54	1.0	. . .
75W-90 synthetic hydrocarbon gear oil, GL-5, MIL-L-2105C, limited slip	-46	-51	28.0	163	325	106	15.2	149	0.56	. . .	. . .
15W-40 motor oil, CE/SG, CD-II, MIL-L-2104E, MIL-L-46152D, Caterpillar TO-2, Allison C-4	-29	-20	28.2	232	450	112	14.8	136	0.53	1.0	7 TBN
15W-40 motor oil, CE/SF, CD-II, MIL-L-2104E, MIL-L-46152D, Caterpillar TO-2, Allison C-3	-31	-24	28.0	226	439	110	14.7	136	0.53	1.5	11 TBN
SAE 40 motor oil, CE/SG, CD-II, MIL-L-2104E, MIL-L-46152D, Caterpillar TO-2, Allison C-4	-23	-9	26.1	243	470	157	15.1	96	0.61	1.0	7 TBN
Polyglycol industrial oil	-30	-22	. . .	280	536	161	22.4	173	0.54	. . .	. . .
SAE 40 motor oil, CC/SF, MIL-L-2104B, MIL-L-46152B	-12	10	26.4	228	442	169	15.5	93	0.62	0.86	6 TBN
ISO VG 220 circulating oil, AGMA R&O No. 4	-9	16	28.7	241	466	138	13.3	89	0.64	. . .	. . .
ISO VG 220 industrial gear oil, AGMA 4EP, U.S. Steel 224	-12	10	26.8	213	415	139	14.6	105	0.60	. . .	. . .
Trunk-type diesel engine oil, CD	15	60	25.6	238	460	140	14.2	92	0.61	3.6	32 TBN
Diesel cylinder oil	-12	10	19.5	243	469	203	17.2	90	0.62	9.3	77 TBN
ISO VG 460 industrial gear oil, AGMA 7EP, U.S. Steel 224	-12	10	25.0	>254	489	480	30.4	90	0.56	. . .	. . .
ISO VG 460 cylinder oil, AGMA 7 comp.	-7	19	24.5	280	536	460	30.5	96	0.55	. . .	. . .

Note: Engine service classifications, such as SG/CD, are defined in Table 8.

Gravity and Density. Gravity is commonly used to identify a particular oil, that is, the density of an oil at 15.6 °C (60 °F) can be calculated by using the density of water as 0.99904 g/cm^3 at 15.6 °C (60 °F). Water has an American Petroleum Institute (API) gravity value of 10. As defined by ASTM D 287 (Ref 6), API gravity is:

$$\text{API gravity} = \frac{141.5}{\text{Specific gravity (15.6/15.6 °C)}} - 131.5 \qquad \textbf{(Eq 1)}$$

Thermal Expansion. Oil density decreases approximately linearly with increasing temperature over a considerable range. Table 4 gives the approximate thermal expansion coefficient for mineral oils (Ref 7).

Table 4 Approximate thermal expansion of mineral oils

Density, g/cm^3 at 15.6 °C (60 °F)	Expansion coefficient, 10^4/K
0.70	11.5
0.75	10.2
0.80	9.1
0.85	8.0
0.90	7.1
0.95	6.3
1.00	5.6
1.05	5.0

Note: Values determined using Roelands equation VIII-9 in Ref 7

Bulk Modulus. Oil density also increases with pressure, and the change in volume is often expressed as bulk modulus (Ref 8). This property, a measure of the resistance to oil compression, is important in the efficiency and dynamic responses of hydraulic and other pressurized-liquid systems.

Isothermal tangent bulk modulus and isentropic bulk modulus are defined thermodynamically, respectively, as $K_t = -V(\partial P/\partial V)_t$ and $K_s = -V(\partial P/\partial V)_s$, where the subscript t refers to a condition of constant temperature and the subscript s refers to entropy. However, the two secant bulk moduli are the ones usually used for engineering purposes.

The isothermal secant bulk modulus is:

$$B_t = \frac{V_0 P}{\Delta V} = \frac{\rho P}{\Delta \rho} \qquad \textbf{(Eq 2)}$$

where V_0 is the volume at atmospheric pressure; P is the gage pressure (0.1 MPa, or 1 atm, equals 0 gage pressure); ΔV is the volume decrease between 0.1 MPa and P; ρ is the density at P; and $\Delta\rho$ is the density increase between 0.1 MPa and P.

The isentropic secant bulk modulus, B_s, is related to the isothermal secant bulk modulus by the ratio of the specific heat at constant pressure, C_p, to that at constant volume, C_V, by the relationship:

$$B_s = \frac{C_P}{C_V}(B_t) \qquad \textbf{(Eq 3)}$$

Wright gives graphical correlations for determining the isothermal secant bulk modulus for mineral oils as a function of temperature and the atmospheric pressure density (Ref 8).

Color is used for lubricant identification purposes, except where staining or appearance are important. In ASTM D 1500, color is matched with standards ranging from the lightest to the darkest, that is, ASTM color 0.5 and 8.0, respectively.

Viscosity, the degree to which a fluid resists flow, is the most important property of a lubricant, and is covered fully in the article "Viscosity and Its Measurement" in this Volume. Viscosity relates directly to the ability of the lubricant to separate bearing surfaces, and often correlates with other performance characteristics.

Viscosity is expressed as being either dynamic (absolute) or kinematic. The latter, measured by ASTM D 445, is commonly utilized by lubricant suppliers and users. Kinematic viscosity is dynamic viscosity divided by density. The conversion factors between current and previously used viscosity units are provided in ASTM D 2161.

Viscosity-Temperature Relation. The viscosity of oils depends strongly on temperature, as shown in Fig. 1. The rate of change with temperature depends primarily on viscosity level, varying from about 0.5%/ °C for a 1 mm^2/s (1 cSt) oil to about 6-10%/ °C for 10^5 mm^2/s (10^5 cSt). The dependence of viscosity on temperature can be represented over a considerable range of temperatures by the ASTM D 341 viscosity-temperature (MacCoull-Walther) equation (Ref 6):

$$\log \log(\text{cSt} + 0.7) = A - B(\log T) \qquad \textbf{(Eq 4)}$$

where A and B are constants and T is absolute temperature in units of Kelvin. The constant 0.7 is valid for viscosities ranging from 2×10^7 down to 2 m^2/s (2×10^7 to 2 cSt); the constant increases as viscosity decreases below this latter value. Additional information regarding the ASTM equation is provided in Ref 9.

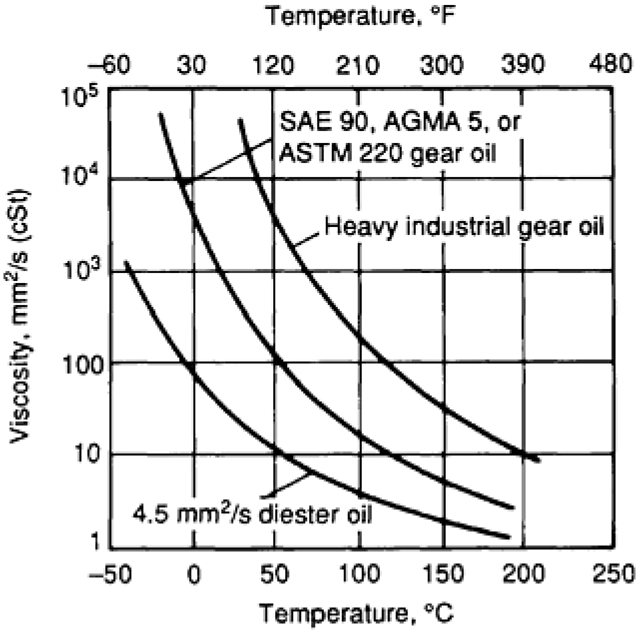


Fig. 1 Comparison of viscosities and temperature dependence

Viscosity index (VI) is an arbitrary method of expressing the viscosity-temperature sensitivity of lubricants, in which high numbers represent smaller changes in viscosity with temperature. Paraffinic oils have high VI numbers, whereas naphthenic oils have low VI numbers. The VI between 0 (the poorest oils in 1929) and 100 (the best in 1929) is given by ANSI/ASTM D 2270 as:

$$\text{VI} = 100\left(\frac{L - U}{L - H}\right) \qquad \textbf{(Eq 5)}$$

where L is the kinematic viscosity of a 0 VI reference oil at 40 °C (105 °F) that has the same kinematic viscosity at 100 °C (212 °F) as the unknown oil at 100 °C (212 °F); H is the kinematic viscosity of a 100 VI reference oil at 40 °C (105 °F) that has the same kinematic viscosity at 100 °C (212 °F) as the unknown oil at 100 °C (212 °F); and U is the kinematic viscosity of the unknown oil at 40 °C (105 °F).

ANSI/ASTM D 2270 tabulates reference oil viscosities and gives a formula for calculating VI numbers over 100.

Viscosity-Pressure Relations. Pressure increases viscosity, and the effects become appreciable for mineral oils at pressures above the order of 10 MPa (1.5 ksi). Classically, the Barus equation expressed the pressure effect

$$\eta_{t,P} = \eta_{t,0} \exp(\alpha P) \quad \textbf{(Eq 6)}$$

where P equals the gage pressure, $\eta_{t,P}$ equals the dynamic viscosity at temperature t and gage pressure P, $\eta_{t,0}$ equals the dynamic viscosity at temperature t and atmospheric pressure, and α equals the viscosity-pressure coefficient.

The Roelands equation is a much improved representation of viscosity change with pressure (Ref 7):

$$\log \eta_{t,P} + 1.200 = (\log \eta_{t,0} + 1.200)\left(1 + \frac{P}{C}\right)^Z \quad \textbf{(Eq 7)}$$

where C is a constant dependent on the units of gage pressure, P (C equals 196.1 MPa, or 28.4 ksi), and Z is a constant characteristic of the liquid under pressure.

For common nonaqueous lubricants, Z generally falls between about 0.5 and 0.8. For mineral oils and other hydrocarbons (including polymers), the following correlation quite accurately estimates Z from dynamic viscosities at 40 and 100 °C (105 and 212 °F) (Ref 7):

$$Z = 7.81\,(H_{40,0} - H_{100,0})^{1.5}\,F_{40,0} \quad \textbf{(Eq 8)}$$

where $F_{40,0}$ equals ($0.885 - 0.864 \times H_{40,0}$) and $H_{t,0}$ equals log [log($\eta_{t,0}$) + 1.200] at temperatures of t °C.

Elastohydrodynamic (EHD) Viscosity-Pressure Coefficient. Simplified hard (high-modulus) EHD equations commonly use the Barus-type α. For actual lubricants with pressure-dependent α, the appropriate value is:

$$\alpha_{EHD} = \left[\int_0^\infty \left(\frac{\eta_{t,0}}{\eta_{t,P}}\right) dP\right] \quad \textbf{(Eq 9)}$$

Figure 2 showns α_{EHD}, calculated with the Roelands equation, as a function of the atmospheric pressure viscosity $\eta_{t,0}$. The figure may be used to determine α_{EHD} for calculation purposes or, for more precise values, Table XIII-2 of Ref 7 can be used.

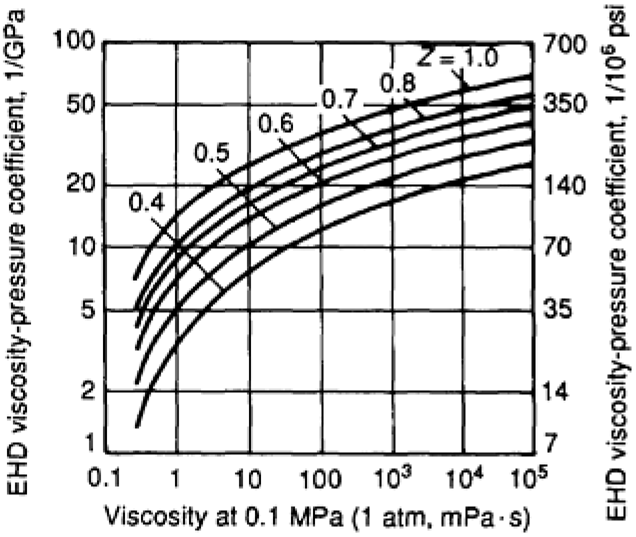


Fig. 2 Elastohydrodynamic viscosity-pressure coefficient. Source: Ref 7

Viscosity at atmospheric pressure and the EHD viscosity-pressure coefficient both strongly depend on lubricant temperature. Hence, for engineering purposes that use simplified EHD equations, it is convenient to express almost all of the lubricant contribution to film thickness in dimensionless terms as follows:

$$\text{Lubricant-speed parameter} = \frac{\alpha_{EHD}\,\eta_{t,0}\,U}{R_x} \quad \textbf{(Eq 10)}$$

where U is the mean surface velocity in the direction of motion and R_x is the effective radius in the direction of motion.

When the lubricant-speed parameter is used in place of the usual speed parameter used by theoreticians, the exponent of their materials parameter, $\alpha_{EHD}E'$ is greatly reduced (E' is the effective elastic modulus of the bearing materials). Hence, effects on the materials parameter by common lubricants only slightly affect calculated film thicknesses. The lubricant portion of the lubricant-speed parameter, $\alpha_{EHD}\eta_{t,0}$, is sometimes called the lubricant parameter.

At the high pressures encountered within the load-carrying zone in hard EHD, lubricants can act as plastic solids with a shear strength. The shear strength increases linearly with pressure according to:

$$\tau_{t,P} = \tau_{t,0} + \beta_t P \quad \textbf{(Eq 11)}$$

where β_t is the pressure coefficient of shear strength at temperature *t;* $\tau_{t,0}$ is the shear strength constant at temperature *t* and 0 gage pressure; and $\tau_{t,P}$ is the shear strength at temperature *t* and gage pressure *P*.

Low-Temperature Flow Properties. At low temperatures, oils become too viscous to pour from a container. Mineral oils also may not pour, because they precipitate crystals of wax at low temperature. Pour point is defined by the ASTM D 97 test.

In addition, the viscosity (that is, the shear stress per unit of shear rate) at low temperatures is commonly not independent of shear rate or stress nor of temperature and shear history, as it is for a "Newtonian" fluid. Therefore, kinematic viscosity (or dynamic viscosity derived from kinematic) determined at low shear rate is no longer a good predictor of how well a lubricant will flow in sumps, pump inlets, oil passageways, and the like. Consequently, a series of tests is used to measure lubricant flow characteristics under conditions that more closely simulate the application shear stresses and shear rates. These methods are:

- ASTM D 2983, Brookfield viscometer
- ASTM D 2602, cold cranking simulator
- ASTM D 4684, borderline pumping temperature of engine oil

Viscosity-Loss. Some lubricants, such as multigraded motor and gear oils, achieve their comparatively high viscosity at elevated temperature without excessive low-temperature viscosity by means of polymeric additives called viscosity index improvers (VIIs). VII oils are non-Newtonian in that the viscosity falls with increasing shear rate. If the shear rate is subsequently reduced after little shearing at high shear rates, then the viscosity returns to its original value and the difference between the low and high shear rate values is termed "temporary viscosity loss." If the oil is sheared extensively at high rates, then mechanical breaking of polymer chains occurs and the viscosity does not return to its original low shear rate level. The loss of low shear rate viscosity is termed "permanent viscosity loss." The temporary viscosity loss of lubricants that have experienced permanent loss is smaller than the temporary loss of a new lubricant.

Temporary viscosity losses typically range between 5 and 30% at the high temperatures of bearings in modern automotive engines (for example, 150 °C, or 300 °F) and shear rates are typically 10^6/s. Permanent viscosity losses also commonly fall in the 5 to 30% range.

A high-temperature, high-shear-rate viscosity of new and sheared oils that correlates with engine performance is measured using the ASTM D 4624 capillary viscometer, the ASTM D 4683 tapered bearing simulator, and the ASTM D 4741 rotational tapered plug tests. Permanent viscosity loss is measured by one or both of the ASTM D 3945 methods for shear stability of polymer-containing fluids using a diesel injector nozzle.

Ash is the mass percent of the oil that remains after combustion. It is used mostly for identification purposes in new oils, but in some cases correlates with deposit and wear performance in engines. Ash is typically measured using the ASTM D 482 method. For unused oils with metal-containing additives, sulfated ash (ASTM D 874) indicates the concentration of the known additive.

Volatility, the tendency to evaporate, is important in terms of fire safety in lubricant handling and use, lubricant consumption under high-temperature and vacuum conditions, and lubricant contamination of the environment. Fire safety is indicated by flash and fire points, which also are sometimes used for oil identification. The flash and fire points are usually determined by the ASTM D 92 Cleveland open cup procedure.

The boiling point range of a lubricant up to 538 °C (1000 °F) is determined by the ASTM D 2887 method using temperature-programmed gas chromatography. The evaporation tendency of lubricants is measured by the ASTM D 2715 procedure, although ASTM D 972 is the most common evaporation test used for motor oils.

Acidity and alkalinity indicate the extent of oxidation of a lubricant and its ability to neutralize acids from exterior sources such as combustion gases. The acidity of lubricants is measured by the amount of potassium hydroxide required for neutralization (mg KOH/g). Basicity is measured in the same units, which is the equivalent of the amount of acid required for neutralization.

Color-indicator methods ASTM D 974 or D 3339 are suitably applied to oils containing acids or bases whose ionization constants in water are greater than 10^{-9}. They are not suitable for many additive oils, especially those containing alkaline detergents, dispersants, or metal-containing inhibitors. For these additive oils, the potentiometric method, ASTM D 664, can be used to determine the total acid number (TAN), strong acid number, total base number (TBN), and strong base number. ASTM D 2896 measures the reserve alkalinity as the TBN, using the potentiometric perchloric acid method.

Stability is discussed below in terms of oxidative and thermal characteristics.

Oxidative stability is the resistance to reaction with oxygen, a natural lubricant "aging" process. Oxidation is undesirable because it increases lubricant viscosity, corrosivity, and deposit-forming tendencies. Oxidation is a sensitive function of time and temperature, oxygen availability, and the presence of water and catalyst metals. It is also sensitive to the mixing and recycling of volatile oxidation products. Oxidation-inhibiting additives can substantially increase the useful life of lubricants, whereas some additives used for other purposes (such as some extreme-pressure additives) can degrade life (Ref 3, 4).

Commonly used test methods are ASTM D 943, D 2272, D 2893, and D 4742. Many proprietary tests are also used. Table 5 shows approximate oxidation-limited temperature ranges as a function of time at temperature for mineral oils with and without oxidation-inhibiting additives.

Table 5 Approximate temperature exposure limits for mineral oils

Exposure time, h	Oxidation				Thermal range	
	Not inhibited		**Inhibited**			
	°C	**°F**	**°C**	**°F**	**°C**	**°F**
1	150-170	300-340	180-193	356-380	410-435	770-815
5	130-152	266-306	163-177	325-350	392-415	738-780
10	122-144	251-290	155-170	310-338	384-407	723-765
50	102-125	216-257	138-153	280-307	365-387	689-729
100	94-118	201-244	130-147	266-297	358-379	676-714
500	74-99	165-210	113-130	235-266	339-359	642-678
1000	66-91	151-196	105-123	221-253	331-351	628-664
5000	46-73	115-163	88-107	190-225	313-331	595-628
10,000	38-65	100-150	80-100	176-212	305-323	581-613

Thermal stability is the resistance of oils to chemical breakdown in the absence of oxygen or water. It can cause carbonaceous or gummy deposits. Thermally induced breakdown is a sensitive function of time at temperature, which, for hydrocarbons, cannot be effectively inhibited. Additives that are used for other purposes in lubricants often are less thermally stable than the base oil. Table 5 also shows the approximate thermally limited temperature ranges for mineral oils.

Carbon residue, which remains after evaporation and pyrolysis, indicates the tendency for coke formation upon the thermal decomposition of ashless oils. In the United States, it is commonly measured by the ASTM D 524 Ramsbottom method. It can also be measured by the ASTM D 187 Conradson method.

Corrosivity is the tendency of a lubricant and its contaminants to chemically react with ferrous and nonferrous metals. Corrosion damages bearings and other structural elements and accelerates lubricant oxidation by catalysis. It is measured in performance tests, including many standard bench oxidation tests. Consequently, oxidation and (nonrust) corrosion properties of a lubricant are commonly considered together. Corrosion can be reduced by additives that inhibit the oxidation process, form protective films on surfaces, or deactivate the catalytic properties of dissolved metals.

Rust consists of hydrated iron oxides and results from aqueous corrosion of ferrous metals. It can damage bearings and interfere with the motion of close-clearance parts, such as hydraulic valves. It also sometimes can breach containment systems and weaken parts. Rust is important because lubricants contain dissolved water and may contain liquid water. In addition, lubricant-wetted parts often are exposed to humid air.

Rust is controlled by using additives that form protective barriers on ferrous surfaces and by reducing the water content of lubricants. ASTM D 665 and D 3603 are commonly used to measure lubricant rust prevention properties.

Detergency and dispersancy are properties that involve the suspension of oil-insoluble materials, in the case of the former, and prevention of sludge and varnish formation, in the case of the latter. The insoluble materials can be oxidation and corrosion products; reaction products of gas-phase materials, such as those that blow by piston rings; or other materials that leak into the lubricant. Both detergency and dispersancy are provided to lubricants by means of additive molecules that consist of insoluble-material-attracting polar groups and oil-attracting groups.

Detergents are oil-soluble salts of organic acids. The base is usually metallic, and typically contains calcium or magnesium. Detergents often contain an excess of alkaline inorganic salts (that is, they are "overbased") that serve to neutralize acids in either blow-by combustion gases or formed by lubricant oxidation. Dispersants are ashless organic compounds that prevent flocculation and coagulation of colloidal particles. The performance of these additives is typically evaluated by a variety of proprietary tests.

Foaming and air release are important properties because machine elements mix air into lubricants. Bubbles that are stable can:

- Reduce heat transfer
- Interfere with lubricant flow
- Cause lubricant to be expelled through vents
- Accelerate oxidation, because of heat generated during compression
- Produce spongy hydraulic-system performance

Foaming is controlled by very low concentrations of antifoam additives. Additives often adversely affect air release. Foaming is measured by ASTM D 892 and other performance-type procedures.

Filterability is the ability to remove particulate matter from lubricants by passing them through porous media. Particles of contaminants cause abrasive wear and may form deposits that interfere with lubricant flow or the motion between parts. Filterability is affected by base oil type and viscosity, additives used for other purposes, and operating conditions. It is determined by a variety of performance tests.

Lubricant Classification

A lubricant can be classified by its viscosity, the type of performance tests it can pass, the type of mechanism for which it is intended, and the industry in which it is used. Lubricants are also classified as automotive, aviation, marine, or industrial lubricants. A particular lubricant generally fits a number of these classifications.

Described below are the most common lubricant categories. Specialized industrial classes, such as paper machine oils, are not included, but can be found in Ref 1, 2, and 3.

Viscosity Grades. Table 6 shows approximate kinematic viscosity levels at 40 °C (105 °F) for several grading systems. The International Organization for Standardization (ISO) viscosity grades (ASTM D 2422) are the nominal kinematic viscosities in mm^2/s at 40 °C (105 °F). They cover the widest viscosity range in increments of about 1.5-fold. The American Gear Manufacturers Association (AGMA) grades are also identified.

Table 6 Comparison of viscosity classifications

ISO viscosity grade, mm^2/s	AGMA number	Approximately equivalent SAE class	
		Engine oil	**Gear oil**
2	. . .	. . .	. . .
3	. . .	. . .	. . .
5	. . .	. . .	. . .
7	. . .	. . .	. . .
10	. . .	. . .	. . .
15	. . .	. . .	. . .
22	. . .	5W	. . .
32	. . .	10W	75W
46	1	15W	. . .
68	2	5W-30, 20-20W	80W
100	3	10W-40, 30	85W
150	4	20W-50, 40	80W-90
220	5	50	90
320	6	60	. . .
460	7	. . .	140
680	8	. . .	. . .
1000	8A	. . .	250

The Society of Automotive Engineers (SAE) grades have viscosity limits set at 100 °C (212 °F), whereas those with W suffixes also have low-temperature requirements. Table 7 shows these limitations for the SAE engine oil classifications (Ref 10). For engine oils, the W requirements enable the starting of engines at cold temperatures and the pumping of enough oil at these temperatures to prevent engine damage.

Table 7 SAE engine oil viscosity classification (J300)

SAE viscosity grade	Viscosity, Pa · s	Borderline pumping temperature (max.)		Stable pour point (max.)		Viscosity, at 100 °C (212 °F), mm^2/s	
		°C	**°F**	**°C**	**°F**	**Min**	**Max.**
0W	3.25[a]	-35	-31	. . .	. . .	3.8	. . .
5W	3.50[b]	-30	-22	-35	-31	3.8	. . .
10W	3.50[c]	-25	-13	-30	-22	4.1	. . .
15W	3.50[d]	-20	-4	. . .	. . .	5.6	. . .
20W	4.50[e]	-15	5	. . .	. . .	5.6	. . .
25W	6.00[f]	-10	14	. . .	. . .	9.3	. . .
20	. . .	. . .	. . .	. . .	. . .	5.6	<9.3
30	. . .	. . .	. . .	. . .	. . .	9.3	<12.5
40	. . .	. . .	. . .	. . .	. . .	12.5	<16.3
50	. . .	. . .	. . .	. . .	. . .	16.3	<21.9

(a) At -30 °C (-22 °F).

(b) At -25 °C (-13 °F).

(c) At -20 °C (-4 °F).

(d) At -15 °C (5 °F).

(e) At -10 °C (14 °F).

(f) At -5 °C (23 °F)

Cross-graded, or multigrade, oils meet the limits of both designated grades (that is, an SAE 10W-30 oil meets the SAE 10W and SAE 30 requirements). Cross grading with mineral-based oils is achieved by using polymeric viscosity-index-improving additives. It also can be achieved by using synthetic-based oils. Outside of the United States, minimum high-temperature, high-shear (HTHS) viscosity (at 150 °C, or 300 °F, and 10^6/s shear rate) is generally specified for the SAE cross grades to protect against wear. Both U.S. and Japanese automobile manufacturers specify a minimum HTHS viscosity for all grades (Ref 11).

Engine oils meet different levels of performance requirements of the American Petroleum Institute (API); the Comité Des Constructeure D'Automobiles Du Marché Commun (CCMC), which is now the Association des Constructeurs Europeens D'Automobiles (ACEA) in Europe; the U.S. military in the United States and Europe; and U.S., European, and Japanese engine builders.

Some premium products meet most of the major requirements of all of these groups. High-quality products generally contain detergent, dispersant, wear inhibitor, friction modifier, oxidation inhibitor, corrosion inhibitor, rust inhibitor, pour depressant, and foam inhibitor additives.

Table 8 gives the current API classifications (arrived at through participation of API, ASTM, and SAE) for gasoline engines, some light-duty diesel engines (S categories), and diesel engines (C categories). There is also an energy-conserving classification. Table 9 shows the engine tests required for the most severe API classifications. Some low-cost oils in the marketplace only meet obsolete classifications, such as SA and SB, which correspond, respectively, to straight mineral oils and oils with modest oxidation and corrosion inhibition (which also can inhibit wear).

Table 8 API engine service classifications

S categories	
SE	Satisfies 1972 U.S. warranty conditions for gasoline engine lubricants. Improved protection against oxidation, high-temperature deposits, rust, and corrosion
SF	Satisfies 1980 U.S. warranty conditions for gasoline engine lubricants. Additives against high- and low-temperature deposits, wear, and corrosion. Improved oxidation stability and antiwear over SE
SG	Satisfies 1989 U.S. warranty requirements for gasoline engine lubricants and meets CC category requirements. Improved antiwear, cleanliness, and antithickening over SF
C categories	
CC	Lightly supercharged diesel engines and certain heavy-duty gasoline engines. Additives against high- and low-temperature deposits, rust, and corrosion
CD	Satisfies requirements of supercharged diesel engines, even with high-sulfur fuels. Additives against high-temperature deposits, wear, corrosion
CD-II	Severe-duty service of supercharged two-stroke engines. Satisfies CD requirements and a supercharged two-stroke multicylinder engine test
CE	Service typical of turbocharged or supercharged heavy-duty diesel engines manufactured since 1983 and operated under both low-speed, high-load and high-speed, high-load conditions
CF-4	Service in high-speed four-stroke-cycle diesel engines, particularly for on-highway heavy-duty truck operations. Exceeds CE requirements and designed to replace CE oils. May be used in place of CD and CC oils. Provides improved control of oil consumption and piston deposits

Table 9 Engine tests for API classification

Gasoline engines
CRC L-38 (CLR engine): Bearing corrosion, oxidation, shear stability
ASTM sequence IID (1977 Oldsmobile V-8 engine): Low temperature, rust, corrosion
ASTM sequence IIIE (1987 Buick V-6 engine): High temperature, wear, and oil thickening
ASTM sequence VE (Ford 4): Low temperature, sludge, varnish, and wear
ASTM sequence VI (1982 Buick V-6 engine): Fuel economy
Diesel engines
CRC L-38: Bearing corrosion, oxidation, shear stability
Caterpillar 1K: Piston deposits
Detroit diesel 6V-92TA (two-stroke engine): Piston deposits, ring and valve distress
Mack T-6: Ring wear, piston deposits, and oil consumption
Mack T-7: Diesel soot dispersion and viscosity increase control
Cummins NTC-400: Piston deposits, bore polishing, and camshaft roller pin wear

Two-cycle oils are used when the lubricant is supplied as a solution in the gasoline fuel or is directly injected in modern engines. Such two-cycle engines are common in boats, snowmobiles, chain saws, lawnmowers, and motorcycles. The oils prevent cylinder wall damage without producing spark plug fouling, surface ignition, or exhaust port plugging. They also provide good rust, corrosion, wear, ring sticking, and varnish protection.

The lubricants are available with diluents to facilitate mixing with gasoline at fuel-to-oil ratios that commonly range between 16 and 100. The engine manufacturer recommends the ratio to be used. Oils are certified by the National Marine Manufacturers Association (NMMA).

Railroad diesel oils are generally either SAE 40 grade or 20W-40 multigrade oils. Typically, they are of API CD quality, but are free of zinc to protect the silver bushings in railroad engines. They have relatively high TBNs to neutralize fuel sulfur acids.

Gas engine oils resist oxidation and nitro-oxidation and are used in engines that burn natural gas or liquified petroleum gas (LPG). They usually are low-ash dispersant-containing oils. However, higher-ash oils are used to neutralize sulfur acids when burning high-sulfur fuels.

Transmission and torque-converter fluids are intended to:

- Transmit power in torque converters and oil-wet clutch packs
- Lubricate the gears, pumps, and splines of transmissions

- Dissipate heat

They have excellent viscosity-temperature characteristics, oxidation resistance, wear prevention, well-controlled dynamic and static friction characteristics, and foam resistance. They also show good seal compatibility. Automatic transmission fluids (ATFs) are classified as Dexron II, Mercon, or Ford type F on the basis of auto manufacturer performance testing. The most common torque fluid classes are Allison C-3, Allison C-4, or Daimler-Benz 236.6.

Gear oil performance classifications range from straight mineral oils to oils compounded with a fatty oiliness additive (for worm gears) or with extreme-pressure (EP) additives (for hypoid gears). Table 10 shows the SAE J308b recommended practice, which covers the API classifications for automotive axles and manual transmissions (Ref 10).

Table 10 API system of lubricant service designations for automotive manual transmissions and axles

API-GL-1	Spiral bevel and worm gear axles and some transmissions under mild service
API-GL-2	Worm gear axles not satisfied by API-GL-1
API-GL-3	Manual transmissions and spiral-bevel axles under moderately severe service
API-GL-4	Hypoid gears in normal severe service without severe shock loading
API-GL-5	Hypoid gears in severest service, including shock loading

AGMA classifications for industrial gearing cover a similar range, with rust and oxidation (R&O), compounded, and EP types. The R&O type provides oxidation and corrosion inhibition and is used for lightly loaded spur and helical gears. The compounded type, with a few percent fatty additive, is for worm gears, whereas the EP type is for hypoid gears and heavily loaded and low-speed spur and helical gears.

Automotive gear oils often have higher EP performance and lower pour points than industrial gear oils. Industrial gear oils often have superior resistance to oxidation and rusting.

API-GL-5 lubricants commonly are qualified under U.S. military specification MIL-L-2105C and, sometimes, under MIL-L-2105D and Mack Truck GO-H. Some API-GL-5 lubricants also provide satisfactory limited-slip differential performance. AGMA EP-type lubricants often meet the U.S. Steel 224 requirement. Open gear lubricants typically contain tackiness additives and may be diluted with solvent for ease of application.

Multiuse lubricants for gears, hydraulic systems, and wet clutches and brakes are commonly used in tractors and other agricultural equipment. These typically meet the performance requirements of one or more manufacturers.

Hydraulic oils are primarily classified in terms of either normal or low flammability. Normal-flammability oils are hydrocarbon based, and range from noninhibited to R&O to antiwear oils. Some are VI improved and others have lubricity additives to prevent friction-induced vibration or noise (that is, stick-slip). Paraffinic mineral oils are most commonly used, but, when low pour points are needed, naphthenic oils are used. Sometimes, synthetic hydrocarbon oils can be used for their low pour point and wide liquid range.

Viscosities at operating temperature and at cold-start temperature are the most important properties (Ref 12). The oils usually are R&O inhibited and sometimes are pour-point depressed. Antiwear, antifoam, and detergent/dispersant additives may be used. Good water separability and filterability are also important properties. High bulk modulus and low gas solubility are desirable for high-pressure systems. Antiwear hydraulic oils protect vane, gear, and certain types of piston pumps.

Fire-resistant oils primarily are phosphoric acid esters. Fire-resistant water-miscible fluids include oil-in-water emulsions, water-in-oil emulsions, solutions of chemicals in water, and water solutions of viscosity-increasing polymeric additives (Ref 12). High-water-based fluids (>90% water) provide good heat transfer, are easily disposable, and the non-oil-containing types are nonflammable in situations where the water cannot be evaporated. However, they are temperature limited, may cause rusting, and require special equipment.

Turbine oils, when premium, have excellent oxidation resistance and water-separation properties. They also have good air separation and rust protection. They use highly refined base oils, mostly paraffinic, but some are naphthenic. Synthetic

hydrocarbon-based fluids are also available for applications requiring exceptional VI and broad liquid ranges. These oils are used for steam turbines, heavy-duty gas turbines, hydraulic systems, and air compressors. They typically satisfy a wide variety of original equipment manufacturer (OEM) and military specifications, as well as meet the AGMA R&O-type gear oil requirements.

Marine steam turbine oils with antiwear additives protect heavily loaded reduction gearing connected to the turbines. These lubricants, otherwise, are premium turbine oils. Phosphate ester lubricants are also used in heavy-duty gas turbines, where their fire resistance is needed.

Aircraft gas turbines are lubricated with synthetic oils that have excellent oxidation and thermal stability. Stationary gas turbines can be lubricated with synthetic or highly stable mineral turbine oils. The lubricants have excellent resistance to deposit formation; good protection against bearing and gear pitting fatigue, as well as corrosion; and good gear load-carrying capacity.

Engine manufacturer and government specifications define the aircraft gas turbine lubricants in three classes. They are often referred to as 3, 5, and 7.5 mm^2/s oils (viscosity at 100 °C, or 212 °F).

Compressor lubricants must be compatible with the gases being compressed, must lubricate, and, in some cases, must seal. The higher the pressure (and, thus, temperature), the greater the tendency of the lubricant to react with the compressed gas and to coke. Reduction in lubricant viscosity by solution of the compressed gas in the lubricant is also a possible compatibility concern (for example, hydrocarbon compression). For air compressors, fire and explosion in the pressurized space and deposit formation are the main concerns. For steam and other wet gas compression, lubricant displacement of water from lubricated surfaces is important.

Mineral compressor oils are usually formulated for other purposes. Premium R&O oils are used for most types of compressors, whereas motor oils are sometimes used for reciprocating trunk-type compressors. Antiwear hydraulic oils, motor oils, or automatic transmission fluids may be required for vane and rotary screw compressors. Cylinder oils used in cross-head compressors can be compounded for wet conditions. Synthetic hydrocarbon-based oils with excellent oxidation stability and good deposit resistance are increasingly used for high-speed, high-temperature compressors. Diester-based oils also are being utilized increasingly for rotary screw compressors.

Refrigerator oils must lubricate the compressor, be thermally stable at compression temperatures (to the order of 160 °C, or 320 °F), be compatible with the refrigerant, and flow at the lowest evaporator temperature (Ref 13). Mineral oil lubricants must be refined to remove the components that can precipitate or react with the refrigerant. Oil entrained in compressed gas is carried through the refrigerant system and must be returned from the evaporator. Thus, miscibility between the oil and refrigerant is important for lubrication and sealing performance. Environmentally safer refrigerants, such as R-134A, require synthetic oils, such as the polyglycol type for miscibility at operating temperatures.

A special test for refrigerator oils is the floc point at which a cooled solution of oil in refrigerant type 12 becomes cloudy, because of precipitate formation. To prevent ice precipitation in refrigerators, oils need to be very dry. The stability of refrigerator oils is typically determined by the amount of deposit formed from a mixture of oil and refrigerant after exposure to metals found in refrigerant systems at elevated temperature (for example, 175 °C, or 350 °F).

Circulation oils are used in systems where oil is circulated to many individual bearings in order to remove large quantities of heat and contaminants. Because good water and air separation, along with good oxidation and rust protection, are often required, R&O oils are most often used. In some applications, straight mineral oils may be satisfactory, whereas other applications may require antiwear protection.

Misting oils are used in mist and fog lubrication systems. They contain polymeric additives to control droplet size so that the oil coalesces on the lubricated part and does not escape as mist.

Health, Safety, and Environment

Lubricant manufacturers are required by law to provide a material safety data sheet (MSDS) for every lubricant in order to satisfy the hazard communication standard of the Occupational Safety and Health Administration (Ref 14). Lubricant suppliers can also furnish information on relevant environmental regulations and laws. However, it is the responsibility of the lubricant user to become familiar with the information and to comply with pertinent regulations. Lubricant manufacturers can provide telephone numbers for medical, safety, transportation, and other emergency assistance.

Toxicity is the ability, upon exposure to a substance, to harm human, animal, or plant life. For lubricants, the usual concern is the effects on humans. Generally, unused lubricants are not highly toxic when exposure occurs through the skin. However, more-toxic contaminants can be accumulated over a period of use. The most common short-term (acute) effect is contact dermatitis, which is a particular problem with cutting oils (which are not otherwise covered in this article). Long-term (chronic) effects, as evaluated by animal tests, indicate carcinogenicity for oils that have not been processed severely enough by solvent extraction or hydrogen treating.

Safety. A MSDS will list the toxic properties in terms of LD_{50}'s (the doses, in mass, of toxic substance per mass of animal that will be toxic to 50% of the animals tested). The prevention of toxic effects on humans requires the avoidance of lubricant contact, including breathing of vapor or mist. Oil-impervious clothing and boots are useful in some circumstances. Thorough washing should follow any personal contact.

The MSDS also lists properties such as flash and fire points. Explosion and fire avoidance measures should be considered whenever a lubricant either becomes hot enough to approach its flash point during normal use or can accidentally contact a flame or hot part, such as when an oil line breaks and sprays oil on an engine exhaust manifold. Flash and fire points can be substantially lowered by lubricant use that provides the opportunity to absorb volatile materials such as gasoline, diesel fuel, or solvents. When a lubricant must be used near or above its flash point, the lubricant/oxidant mixture must be kept either too lean or too rich to burn. It is good practice to read and understand the precautionary labels on a lubricant container, as well as the MSDS.

Environmental protection requires elimination of lubricant escape to air, water, or land. This entails careful storage and handling of both new and used lubricants, lubrication procedures, equipment maintenance (especially seals, gaskets, valves, and fittings), and disposal of used lubricants. It also entails avoidance of accidental release, measures that minimize the impact of a release that does occur, and plans to remediate any impact.

Lubricant disposal is costly and subject to evolving federal, state, and local regulations. Improper disposal is a potentially expensive future liability. Consequently, the minimization of used or leaked lubricant disposal is commonly cost effective (Ref 15). Used lubricant minimization involves engineering to reduce aging and other contamination of the oil, system maintenance, and periodic oil testing to determine used oil condition.

When analysis indicates that a lubricant is no longer suitable for service, it can often be reconditioned for further use, either on-site or off-site, by a contract recycler. Such recycling commonly involves water removal by gravity, centrifuge, coalescer, or vacuum evaporation, and fine-particle filtration. It also may involve clay treatment and possible additive refortification. Portable water removal and filtration units are often used at sites that have a number of lubricant systems.

Disposal is required for lubricants that can no longer be reconditioned. Over half of the used oil in the United States is utilized as fuel. It can be burned in industrial furnaces if contaminant concentrations do not exceed the limits for arsenic (<5 parts per million by mass, or ppmm), cadmium (<2 ppmm), chromium (<10 ppmm), lead (<100 ppmm), and total halogens (<4000 ppmm); if the flash point is at least 37.8 °C (100 °F) (Ref 16); and if it does not contain toxic substances. About one-third of the used oil is dumped and small percentages are rerefined, used for other industrial uses, or used for road oiling.

In view of the changing regulations and potential liabilities, good written records should document the source of the waste oil and its subsequent handling (storage, transportation, and disposal). An analysis of the used oil is desirable, and a retained sample also may prove useful in establishing that the oil was not contaminated. The generator of the used oil should contract with a waste oil hauler who carries adequate insurance and has a licensed treatment, storage, and disposal facility that complies with all federal and state regulations.

Methods of Lubricant Application

The lubricant application method plays a vital role in how the lubricant functions. The quantity of lubricant, its temperature, and its cleanliness are as important to bearing system performance as the selection of the proper lubricant.

Methods of providing lubricant to a bearing range from periodic manual application with a traditional oil squirt can to continuous automatic metering from a circulating oil system supplying an entire machine or group of machines (Ref 1, 2, 5). The appropriate method should supply the proper quantity of oil at a correct rate. Considerations that are involved in selection are whether the supply needs to be continuous, its adaptability to changed operating conditions, and the reliability of the method. Reliability considerations include the human factors and the effects of factors such as

temperature, sump level, and oil contamination on the quantity of oil delivered. In addition, the economics of the method and its convenience, safety, and environmental compatibility need to be considered.

Manual Application. The oil squirt can, a sprayer, or a brush is often satisfactory for machine elements used only occasionally, for low-speed lightly loaded bearings, or for inexpensive or rough machinery, such as wire rope, chains, or open gears. The oiling is intermitted and depends on periodic human action to supply the proper lubricant quality and quantity.

Drop oilers provide lubricant one drop at a time. Gravity-fed systems are common, and shut-off can be either manual or automatic. They are most economical to use when there are relatively few easily accessible lubrication points. Oiling is intermittent, and the oilers usually need to be manually resupplied periodically.

Splash lubrication is provided automatically and continuously to an enclosed mechanism by immersing the lower portion of a rotating or vertically reciprocating part in an oil sump.

Oil Carrier. The self-acting arrangements that can be used to carry oil from a reservoir to a bearing are described below.

Wick and Pad. A fibrous material carries oil by capillary action. Wicks can clog with an accumulation of oil contaminants. Pads of fibrous or otherwise open-pored materials that are filled with oil and are in contact with a bearing can also be used to apply lubricant. Accumulation of dirt, wear particles, and lubricant degradation products limit pad life.

Oil rings or chains are used for horizontal bearings with rotating shafts. They rest on the shaft and rotate with it, because of friction. The lower part dips into oil in a sump below the bearing and automatically drags oil by viscous forces as the oil-wet lower part rotates to the top of the shaft. The chain oiler has greater lubricant-carrying capacity than a ring oiler, but is limited to lower speeds by centrifugal action and churning drag.

An oil collar acts similarly to an oil ring, but is fixed to the shaft. It carries oil from the sump and the oil is displaced from the outer portions of the collar to the shaft by means of scrapers.

Pressurized Feed. Pumping lubricant directly to individual bearings provides positive oil feed to each. It is easily made automatic and is readily adaptable to recirculation of the supplied oil for filtration, temperature control, and other purposes. Usually, multiple bearings on a machine or mill are lubricated by the same oil-circulating system.

Pressurized feed circulating systems have the disadvantage of being initially expensive. However, their ability to provide dependable continuous flow of conditioned oil makes them attractive for applications involving heavy-duty expensive bearings.

Mist lubrication is provided by a mist of oil droplets in air that impinge on bearings. These once-through systems consist of an oil mist generator, plumbing, and droplet-sized reclassification nozzles. They can be designed to deliver oil automatically to many bearings spread over distances up to about 100 m (330 ft). Mist lubrication reduces energy losses that result from excess lubricant in bearings.

These systems commonly cost less initially than pressure-fed circulating systems. Although they are very dependable, the escape of stray mist into surrounding air can be an environmental problem, and deposition of oil degradation products can progressively clog mist fittings. Mist systems cannot remove an appreciable amount of heat from a bearing, compared with a pressure-fed circulating system.

References

1. E.R. Booser, Ed., *CRC Handbook of Lubrication: Applications and Maintenance,* Vol 1, CRC Press, 1983
2. E.R Booser, Ed., *CRC Handbook of Lubrication: Theory and Design,* Vol 2, CRC Press, 1983
3. D. Klamann, *Lubricants and Related Products,* Verlag Chemie, Federal Republic of Germany, 1984
4. M.J. Neale, *Tribology Handbook,* Wiley, 1973
5. J.J. O'Connor and J. Boyd, *Standard Handbook of Lubrication Engineering,* McGraw-Hill, 1968
6. *1990 Annual Book of ASTM Standards,* ASTM, 1990

7. C.J.A. Roelands, *Correlational Aspects of the Viscosity-Temperature-Pressure-Relationship of Lubricating Oils,* Doctoral thesis, Technische Hogeschool te Delft, Netherlands, 1966
8. W.A. Wright, Prediction of Bulk Moduli and Pressure-Volume-Temperature Data for Petroleum Oils, *ASLE Trans.,* Vol 10, 1967, p 349-356
9. E.E. Klaus and E.J. Tewksbury, *CRC Handbook of Lubrication: Theory and Design,* Vol II, E.R. Booser, Ed., CRC Press, 1983
10. *1991 SAE Handbook,* Society of Automotive Engineers, 1991
11. *The ILSAC Minimum Performance Standard for Passenger Car Engine Oils,* adopted by Japan Automobile Manufacturers Association and Motor Vehicle Manufacturers Association of the United States, Oct 1990
12. D. Klamann, *Lubricants and Related Products,* Section 11.9, Verlag Chemie, Federal Republic of Germany, 1984
13. D. Klamann, *Lubricants and Related Products,* Section 11.7.2, Verlag Chemie, Federal Republic of Germany, 1984
14. *Code of Federal Regulations,* Vol 29, Office of the Federal Register, National Archives and Records Administration, Part 1910.1200
15. T.L. Lantz, Lubricant Conservation, *Lubr. Eng.,* Vol 44 (No. 5), 1988, p 408-411
16. *Code of Federal Regulations,* Vol 40, Office of the Federal Register, National Archives and Records Administration, Part 266.40

Lubrication Regimes

Herbert S. Cheng, Northwestern University

Introduction

WHEN TWO SURFACES are in contact with each other, the load is carried by many high points, or asperities, on the surfaces. During sliding, the total tangential force required to shear these asperity junctions is usually high, causing unacceptable friction, wear, and surface damage. To reduce the frictional force and thus allow easier sliding, a lubricant is deliberately introduced to separate the asperities either totally or partially.

The use of liquid or gas lubricants is known as *fluid-film lubrication. Thick-film lubrication* refers to the total separation of asperities by a lubricant film thickness many times larger than the size of the lubricant molecules. If this condition exists only partially--that is, if part of the load is carried by the fluid pressure and the rest is borne by contacting asperities separated by a molecularly thin lubricant film--the term *thin-film lubrication* or sometimes *mixed lubrication* is used. In the most severe form of thin-film lubrication, the entire load is carried by asperities lubricated by surface films of molecularly thin liquids, gases, or solids; this condition is known as *boundary lubrication.* The exclusive use of solid lubricants is called *solid lubrication.*

The lubrication between two sliding surfaces can shift from one of the three regimes--thick-film, thin-film, or boundary lubrication--to another, depending on the load, speed, lubricant viscosity, contact geometry, and surface roughness of both surfaces. This dependence was first recognized in 1902 by Stribeck, who observed the variation of the sliding friction with a lubrication parameter $\eta N/p$, where η is the lubricant viscosity, N is the angular velocity of the cylindrical contact, and p is the average contact pressure. Figure 1 shows a typical Stribeck curve. At the right sides, where the friction increases slightly with the lubrication parameter, lubrication is in the thick-film regime; at the far left, where the friction is nearly constant, lubrication is in the boundary regime. In the middle, lubrication is in the mixed-mode, or thin-film regime. The boundaries of these regimes would move to the right if the surfaces became rougher and to the left if they became smoother. In addition to the variables η, N, and p, other parameters related to the study of lubrication regimes are defined in Table 1.

Table 1 Nomenclature for calculating regimes of lubrication

Symbol	Definition	Comments
G	Dimensionless material parameter	. . .
$G\infty$	Limiting shear modulus	. . .
h_0	Inlet film thickness	. . .
h_{min}	Minimum film thickness	. . .
h^*	Asperity film thickness or microelastohydrodynamic film thickness	Thickness of the oil film drawn into the asperity contact by microelastohydrodynamics; also referred to as the micro-EHL film thickness. First line of defense against sliding failure in mixed lubrication
$\bar{h}$	Average lubricant film thickness	Measure of the effectiveness of hydrodynamic lubrication. Depends on the speed, viscosity, pressure-viscosity exponents, contact radii, and the orientation of the roughness lays with respect to the entraining velocity. Theories in mixed lubrication are reasonably well developed (Ref 1, 2) to predict $\bar{h}$ as affected by the surface roughness height and lay orientation.
N	Angular velocity of cylindrical contact	. . .
p_i	Pocket pressure in hydrostatic bearing	. . .
p^*	Local asperity contact pressure	Represents the maximum contact pressure above the surrounding fluid pressure at each asperity contact. Its value depends on the height and the slope of asperities. Because the height and the slope are random functions, p^* is also a random function. The distribution of p^* controls the asperity shear stress and asperity contact temperature.
$\bar{p}$	Average (bulk) hydrodynamic pressure	From the elastohydrodynamic action, the lubricant pressure is generated within the Hertzian conjunction. For thick-film elastohydrodynamic lubrication, the distribution of pressure is a smooth function. However, for thin-film EHL, the pressure distribution is not smooth and contains fluctuations at the asperity contacts. $\bar{p}$ can be used as the average lubricant pressure, which varies from point to point within the Hertzian conjunction but ignores the local pressure fluctuations around the asperities. The average lubricant pressure in concentrated contacts can be predicted by lubrication analysis for rough surfaces (Ref 3).
$\bar{p}_a$	Average (bulk) asperity contact pressure	At the asperity contacts, the local pressure is intensified due to asperity deformation. These asperity contact pressures can be average out over a small area within the Hertzian conjunction to form a smooth function for the average contact pressure, $\bar{p}_a$. The average asperity contact pressure in concentrated lubricated contacts can be predicted from Patir and Cheng's analysis (Ref 3) based on the load and compliance relation developed by Greenwood and Tripp (Ref 4).
R	Radius of equivalent cylinder	. . .
$\bar{T}_s$	Average (bulk) surface temperature rise	Average surface temperature rise generated by the fluid shearing and the sliding asperities
T_s	Local asperity contact temperature rise	Indicates the temperature rise above the surrounding surface temperature at each sliding asperity. It is also a random function.
U	Dimensionless speed parameter	. . .
W	Dimensionless load parameter	. . .
γ	Pressure viscosity coefficient	. . .
η	Lubricant viscosity	. . .
η_0	Ambient viscosity of lubricant	. . .
θ	Angle of bite	. . .
Λ	Film thickness parameter	. . .
σ	Workpiece flow strength	. . .
τ_L	Limiting shear stress	. . .
τ^*	Local asperity contact shear stress	Represents the maximum shear stress at a sliding asperity above the surrounding fluid shear stress
$\bar{\tau}$	Average lubricant shear stress	Represents the average shear stress in shearing the thin lubricant film within the Hertzian conjunction

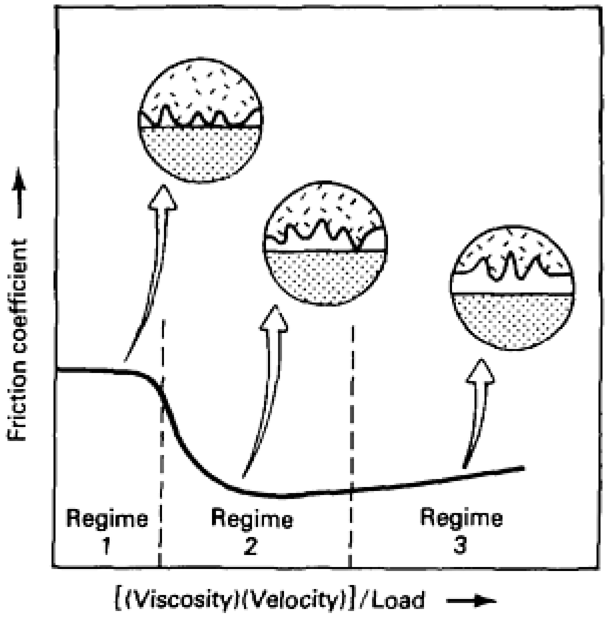


Fig. 1 Plot of friction coefficient, μ versus [(viscosity) (velocity)]/load, ($\eta_0 U$)/*P*, to show range of the three regimes of lubrication. Regime 1, boundary lubrication; regime 2, thin-film lubrication; regime 3, thick-film lubrication

Thick-Film Lubrication

A thick lubricant film can be generated by the tangential and normal relative motion between two surfaces. This mode is known as hydrodynamic lubrication. The effectiveness of such lubrication depends directly on relative speed and lubricant viscosity. For very slow contacts, a thick film is not likely to develop unless an externally pressurized lubricant is introduced into the lubricant film. This type of lubrication is known as hydrostatic lubrication.

For highly loaded contacts, elastic deformation of the surfaces can redistribute and broaden the contact area and lubricant pressure, thus greatly increasing the load capacity and lubricant film thickness compared with those generated by rigid contacts. Such thick-film lubrication resulting from the surface flattening effects of elastic deformation is known as elastohydrodynamic lubrication (EHL). In many metalforming operations, a thick lubricant film can also be formed at the interface between the rigid die surface and the plastically deformed surface of the workpiece. This mode of thick-film lubrication is often referred to as plastohydrodynamic lubrication (PHL).

Hydrodynamic Lubrication

When fluid lubricant is present between two rolling and/or sliding surfaces, a thick pressurized film can be generated by the surface velocities to reduce friction and wear. This mode of lubrication is commonly called hydrodynamic lubrication. Hydrodynamic film thickness can be formed by wedging the lubricant through a convergent gap with the tangential surface velocities, known as wedging film action, or by squeezing the lubricant out of the contact area with the relative normal velocity between the contacting surfaces, known as squeeze film action.

Wedging Film Action. In a converging hydrodynamic slider (Fig. 2), the thicker film at the inlet section can transport more lubricant than the thinner film at the exit section. Because the flow in an infinitely wide slider must be constant throughout the entire section, the lubricant pressure must rise at the inlet section to impede the flow and decrease at the exit section to enhance the flow, as shown in the pressure profile in Fig. 2. The pressure generated in such a converging gap represents the basic wedging film action in hydrodynamic lubrication.

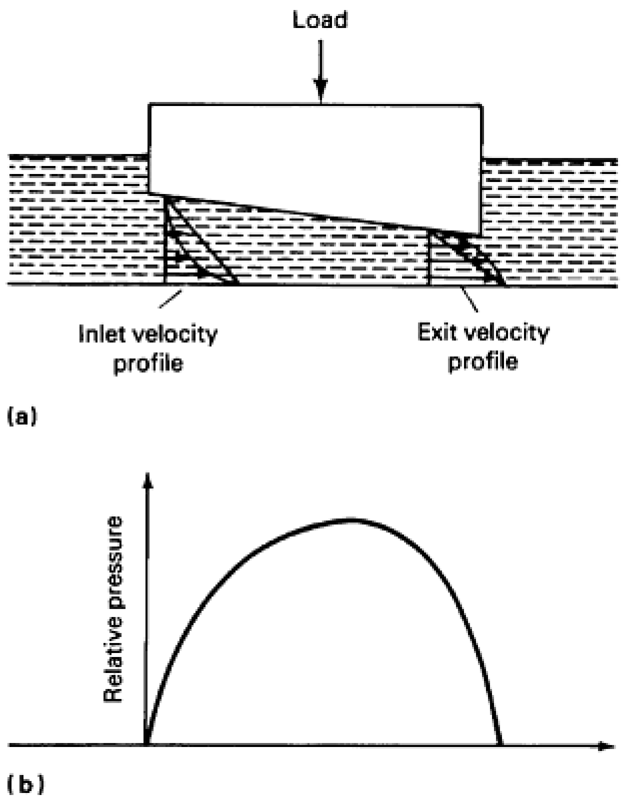


Fig. 2 Wedging film action in a hydrodynamic slider. (a) Velocity profiles at inlet and exit regions of wedge-shaped load. (b) Pressure distribution beneath wedge

Wedging film action takes place not only in flat sliders but also in curved sliders, between cylindrical surfaces such as journal bearings and lubricated rollers, and in spherical surfaces such as ball joints and ball bearings. In these cases, the gap profile may contain a divergent section in the exit section. Under such conditions, the pressure may terminate not at the end of the exit section but rather somewhere in the divergent section because of cavitation.

Squeeze Film Action. For a flat slider with a perfectly parallel gap, hydrodynamic pressure will not be generated by the wedging action. However, if the slider moves downward at a velocity normal to the surface, lubricant pressure will be generated by squeezing out the lubricant at both edges. For a parallel film, a parabolic pressure profile will be generated (Fig. 3).

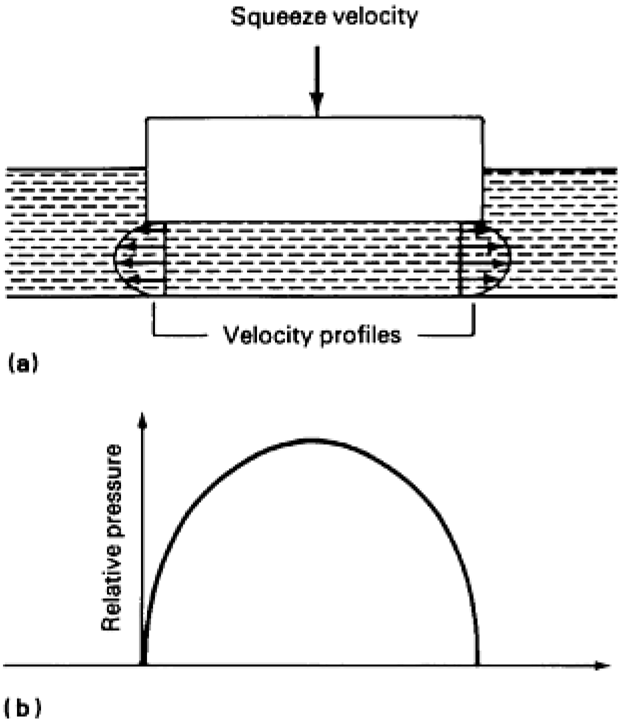


Fig. 3 Squeeze film action in a hydrodynamic slider. (a) Equal velocity profiles generated at each end of a flat slider that moves downward due to a squeeze velocity directed normal to the slider surface. (b) Pressure distribution beneath flat slider

For bearings operating under dynamic loads or reciprocating motion, squeeze film action can develop along with wedging action. The pressuregenerated by the squeeze action can become significant and often provides an effective damping component for stabilizing high-speed rotor-bearing systems.

Lubricant film thickness is important in hydrodynamic lubrication for designing against possible solid-to-solid contact. It is usually determined from solutions of the flow continuity equation, which calculates the lubricant pressure for a known film thickness. Charts and computer software are available for determining the lubricant film thickness for many common bearing geometrical configurations. Lubricant film thickness increases with the sum of the two surface velocities, lubricant viscosity, and bearing size, and decreases with load.

Figure 4 shows the Raimondi-Boyd design chart (Ref 5) for determining the minimum film thickness and eccentricity ratio for a 360° journal bearing. Design chart parameters are defined in Table 2. The effects of journal speed, viscosity, load, and the clearance-to-radius ratio are combined in a dimensionless Sommerfeld number, *S*. Charts for calculating minimum film thickness for other types of journal bearings can be found in Ref 5. The minimum film thicknesses for sliders used for thrust bearings can be calculated using methods outlined by Arnell *et al.* (Ref 6).

Table 2 Nomenclature for Raimondi-Boyd design chart in Fig. 4

Symbol	Definition
c	Radial clearance of bearing
d	Bearing diameter
f_{min}	Minimum friction force
h_0	Minimum film thickness: $h_0 = c(1 - \epsilon)$
l	Bearing length
N	Bearing speed (rev/s)
P	Load per unit projected bearing area
r	Bearing radius
S	Sommerfeld number: $S = \left(\frac{r}{c}\right)^2 \frac{\mu N}{P}$
W_{max}	Maximum load
ϵ	Eccentricity ratio: $\epsilon = [1 - (h_0/c)]$
μ	Effective lubricant viscosity

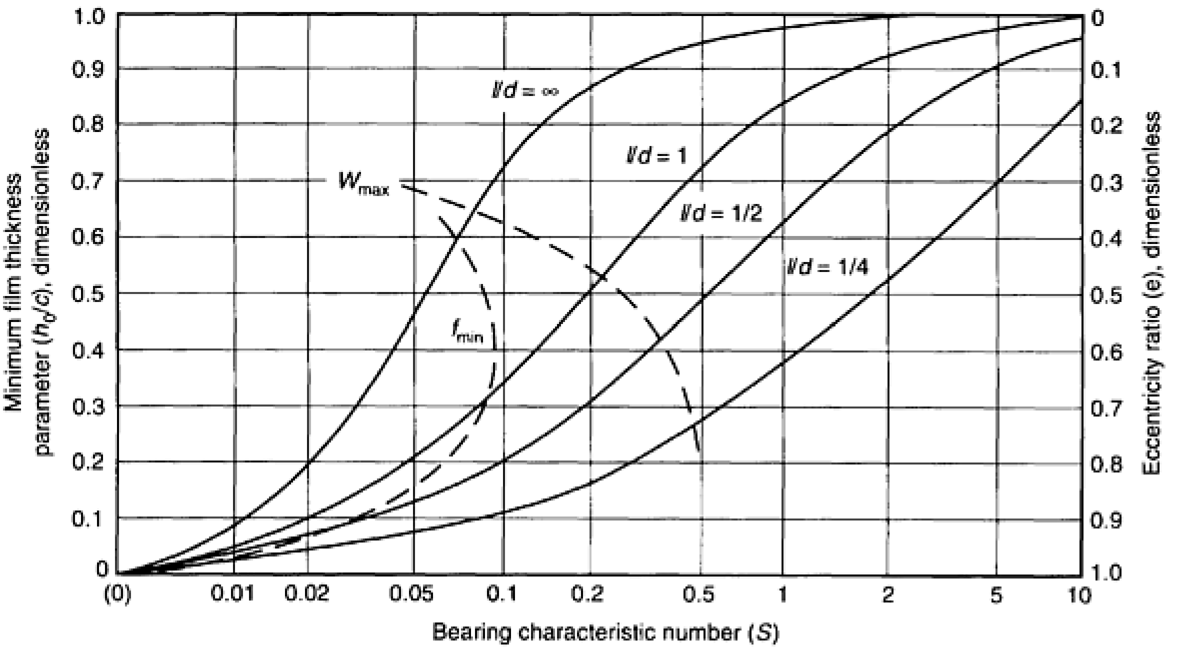


Fig. 4 Raimondi-Boyd design chart used to obtain the minimum film thickness and eccentricity ratio for a 360° hydrodynamic journal bearing. See Table 2 for definition of parameters.

Thermal Effects. Temperature significantly affects hydrodynamic lubrication by reducing lubricant viscosity and film thickness. For this reason, lubricants with a high thermal stability that yields a lower reduction in viscosity at elevated temperatures are generally preferred.

Determinations of lubricant film thickness that fully account for thermal effects require very complex analyses and lengthy computations. In most cases, hydrodynamic bearing lubrication designs including thermal effects can be based on an effective viscosity derived from a gross heat balance analysis and a film thickness calculated using isothermal analysis. These methods are described in Ref 5 and 6.

High-Speed Effects. For hydrodynamic bearings operating at high speeds, isothermal and gross heat balance analyses may not be sufficiently accurate for determining the regime of lubrication. Instead, a full thermal analysis for calculating the lubricant film thickness is usually required.

An increase in lubricant temperature reduces viscosity and yields a film thickness lower than that given by isothermal analysis. Because the reduction in viscosity is higher than the reduction in film thickness, friction normally decreases slightly when lubricant temperature is increased at high speeds.

In addition to the thermal effects, lubricant flow may become turbulent at high speeds because of the reduction in viscosity. The turbulence increases the flow resistance and the effective viscosity. This in turn generates a higher friction, a higher film temperature, and a slightly larger film thickness.

In many high-speed steam turbine compressor pumps, the fluid-film bearings operate in the regime where thermal and turbulence effects are both significant. To ensure that operation lies in the regime of full-film hydrodynamic lubrication, thermal hydrodynamic analysis using an effective turbulent viscosity, such as that contributed by Satar and Szeri (Ref 7) is often needed to determine an accurate film thickness.

Dynamic Loads. In many hydrodynamic bearings, the load is not steady. It changes because of variable external loading, such as the gas load during the firing cycle of a combustion engine or the unbalanced inertia forces of a high-speed rotor or reciprocating piston. In such cases, the lubricant film thickness is also unsteady and fluctuates periodically in response to the imposed periodic load. The ratio of the minimum film thickness to roughness in the fluctuating cycle calculated by transient film analysis indicates whether the bearing is operating in the full or partial hydrodynamic regime.

Transient film analysis includes both the wedging and squeeze film actions. The squeeze film action provides a cushion at the peak load and generally yields a larger minimum film thickness than the static analysis, which neglects the squeeze film action. Typical examples of dynamically loaded bearings include the main bearings and camshaft bearings in combustion engines and high-speed turbine bearings supporting dynamic loads caused by rotor unbalances.

Reciprocating Motion. In reciprocating sliders and oscillatory journal bearings, the cyclic sliding velocity also causes the film thickness to be cyclic, thus producing a squeeze film action. Typical examples include piston rings, piston skirts, and wrist pin bearings. In these cases, the minimum film thickness occurs at or near the top or bottom dead center of the reciprocating motion and can be determined by transient film analysis.

Hydrostatic Lubrication

If the surface velocities are insufficient to generate a thick film in hydrodynamic lubrication, hydrostatic lubrication, which uses an externally pressurized lubricant to generate a thick film, is often employed. Hydrostatic bearings are generally used in very low-speed applications such as machine tool guideways and radar antenna supporting bearings, in extremely low-friction devices such as instrument bearings, in cases of low-viscosity lubrication such as water and air bearings and in such applications as lifting a heavy rotor during startup and suppressing the rotor bearing instability in high-speed hybrid (that is, hydrodynamic and hydrostatic bearings).

Figure 5 shows the basic configuration of a hydrostatic bearings. The externally pressurized lubricant is first fed through a restrictor into a central pocket and then leaks through the bearing area to the outside. The pressure in the pocket (p_i), the film thickness across the bearing area (h), and the lubricant flow depend on the bearing load. For a heavy load, the pocket pressure approaches the supply pressure, yielding a very small flow and thin film thickness. For a light load, the pocket pressure becomes small, producing a high flow and a large gap.

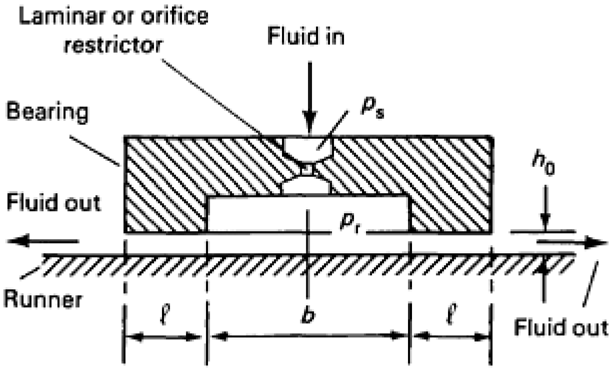


Fig. 5 Schematic showing key parameters that determine operation of a hydrostatic bearing. Nomenclature: p_s, supply pressure; p_r, recess pressure; h_0, film thickness; b, bearing pocket diameter; ℓ, bearing load thickness

Variations of load, flow, and stiffness with the lubricant film thickness of a hydrostatic bearing differ, depending on the types of restrictors used in the bearing. Figure 6 presents typical curves that show differences between the dimensionless loads, flows, and stiffnesses for orifice and laminar restrictors. Table 3 defines the terms used in Fig. 6. These curves can be used as a guide for selecting a restrictor geometry that will yield a film thickness satisfying the following requirements:

- The film must be sufficiently thick to avoid asperity contact
- Combined frictional and pumping loss must be low
- Leakage rate must be less than the maximum tolerable
- Stiffness must be high to prevent a large excursion of film thickness under fluctuating loading

Table 3 Nomenclature for hydrostatic bearings with orifice or capillary restrictor as plotted in Fig. 6

Symbol	Definition
b	Bearing pocket diameter
h	Film thickness across bearing area
h_0	Design lubricant film thickness
$\bar{h}$	Average lubricant film thickness: $\bar{h} = h/h_0$
L	Length of recess
p_r	Recess pressure
p_s	Supply pressure
$\bar{p}$	Average pressure
Q	Lubricant flow
$\bar{Q}$	Average lubricant flow: $\bar{Q} = Q\left(\frac{h_0^3 L p_s}{6 \mu \ell}\right)$
r	$r = p_r/p_s$
S	Stiffness
$\bar{S}$	Average stiffness: $\bar{S} = S\Big/\left[\frac{p_s L(b+\ell)}{h_0}\right]$
W	Load
$\bar{W}$	Average load: $W/[p_s L(b+\ell)]$
μ	Viscosity
ℓ	Bearing land thickness

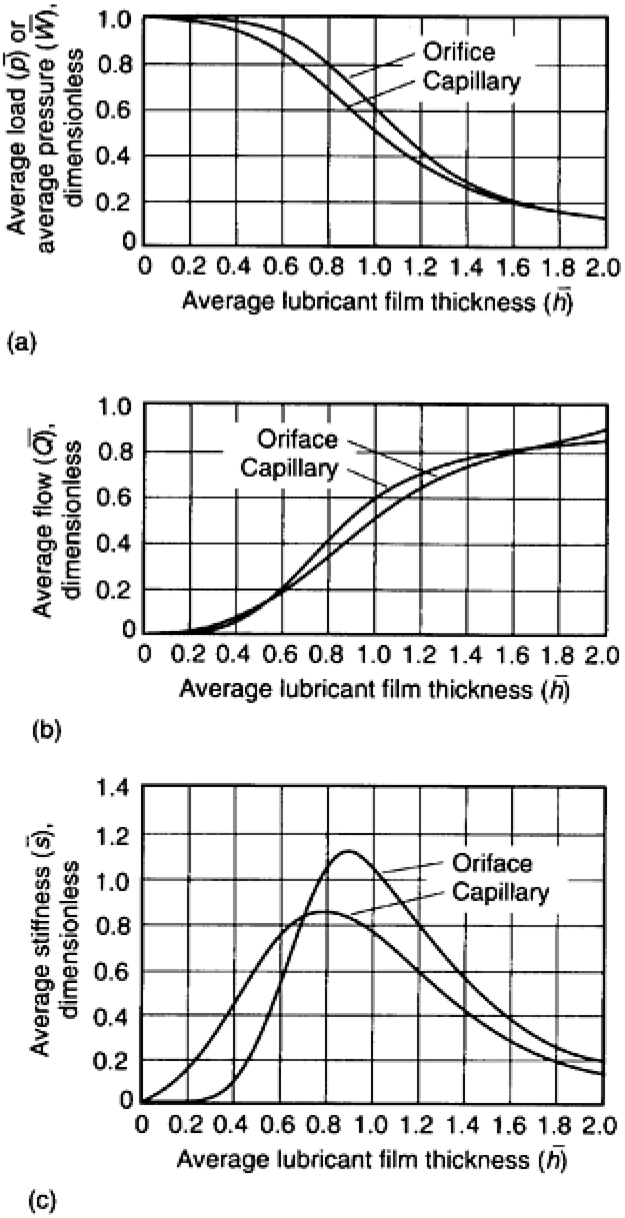


Fig. 6 Selected properties of hydrostatic bearings, which incorporate an orifice or a capillary restrictor, as a function of average lubricant film thickness. (a) Load capacity. (b) Flow rate. (c) Stiffness. Orifice parameters: *r* = 0.586 at $\bar{h}$ = 1.0. Capillary parameters: *r* = 0.500 at $\bar{h}$ = 1.0

Elastohydrodynamic Lubrication

When lubricant pressure causes elastic deformation of the surfaces that is on the same order as the lubricant film thickness, the influence of deformation on lubrication performance becomes a significant parameter. Contacts operating under this condition are in the regime of elastohydrodynamic lubrication.

A majority of lubricated contacts, such as rolling bearings and gear teeth, have surface deformation comparable to or exceeding the lubricant film thickness. Therefore, EHL is extremely important in determining friction and wear in many mechanical components. Characteristics of EHL in the thick-film regime are reasonably well understood and are reviewed in the following sections.

Film Thickness and Shape. In EHL, lubricant film thickness is the most important variable. It is formed by the wedging action at the inlet region between an elastically deformed Hertzian contact under a steady load. As in hydrodynamic lubrication, it increases with rolling speed, lubricant viscosity, and roller or ball size, and decreases with load. The influence of each of these effects varies, depending on contact pressure and modulus. For high-modulus

contacts, commonly known as hard EHL contacts, the film thickness also increases with the lubricant pressure-viscosity dependence. This effect is absent in low-modulus, or soft EHL, contacts. Examples of hard EHL contacts include metallic gears, rolling bearings, cams, and other Hertzian contacts. Examples of soft EHL contacts include all compliant bearings, elastomeric seals, and artificial hip and knee joints.

For hard EHL roller contacts, the film distribution as measured by interferometry (Ref 8) in the conjunction zone (Fig. 7) is largely uniform, with a small constriction at the exit of the conjunction. The minimum film thickness, h_{min}, for hard EHL contacts can be predicted using the Dowson and Higginson equation (Ref 9):

$$\frac{h_{min}}{R} = 2.65 \frac{G^{0.54}U^{0.7}}{W^{0.13}} \quad \textbf{(Eq 1)}$$

Equation 1 applies for moderate rolling speeds. For high-speed contacts, the loss of viscosity due to inlet heating will reduce the film thickness. This reduction can be determined by a thermal reduction factor based on numerical results obtained by Murch and Wilson (Ref 10).

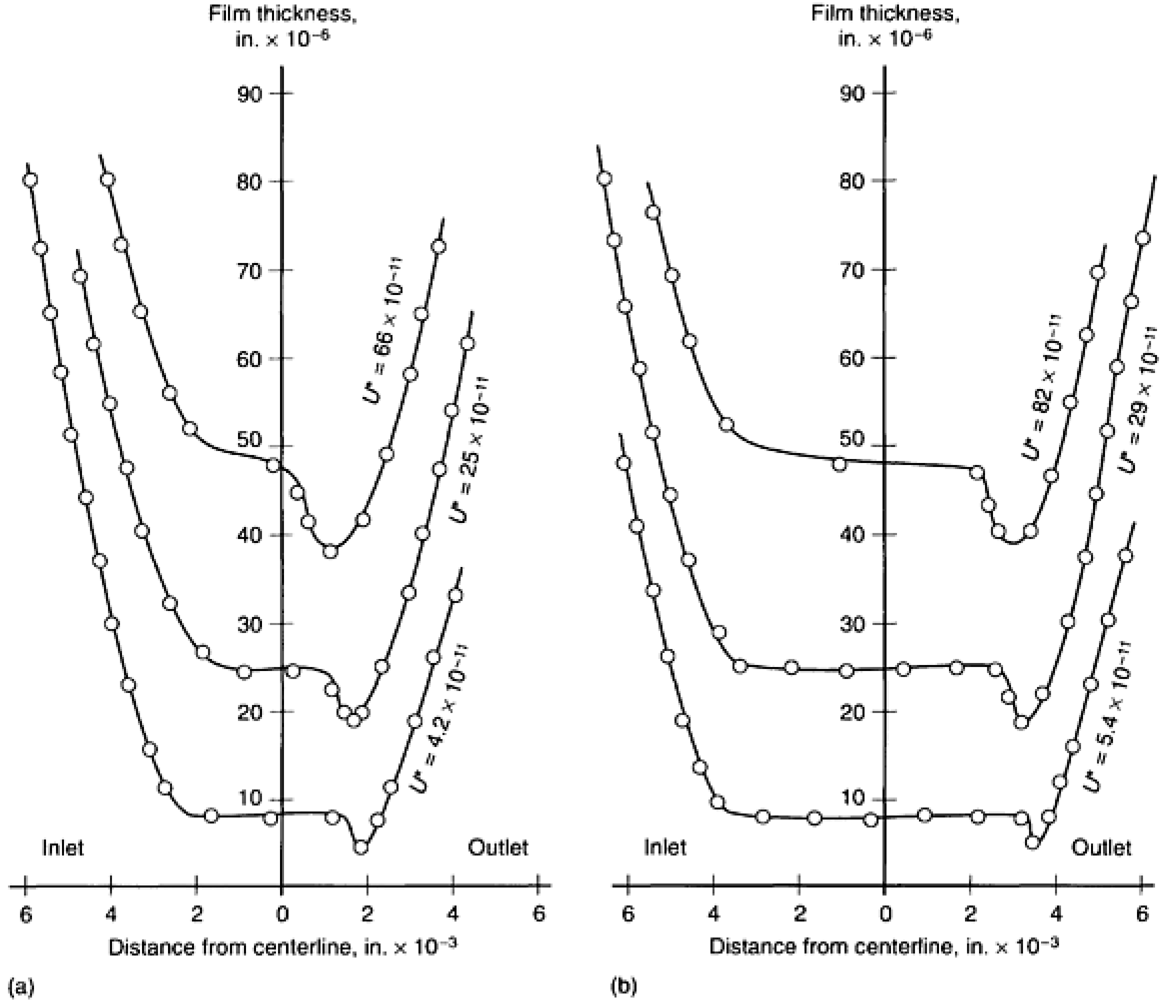


Fig. 7 Film profiles of EHL hard roller contacts at the midpoint of the conjunction zone in the direction of rolling for selected loads, *W*. (a) *W*, 750 N (168.5 lbf). (b) *W*, 1550 N (349 lbf). *U, a dimensionless speed diameter**

For elliptical conjunctions with the major axis oriented perpendicular to the rolling direction, the film thickness is still largely uniform over the conjunction zone. However, the uniform thickness is reduced slightly around the trailing edge, as shown in Fig. 8. The reduction along the two sides is greater than at the center. Empirical formulas for film thickness

based on extensive numerical analyses were developed by Hamrock and Dowson (Ref 11, 12, 13, 14) for the nominal and side minimum thicknesses.

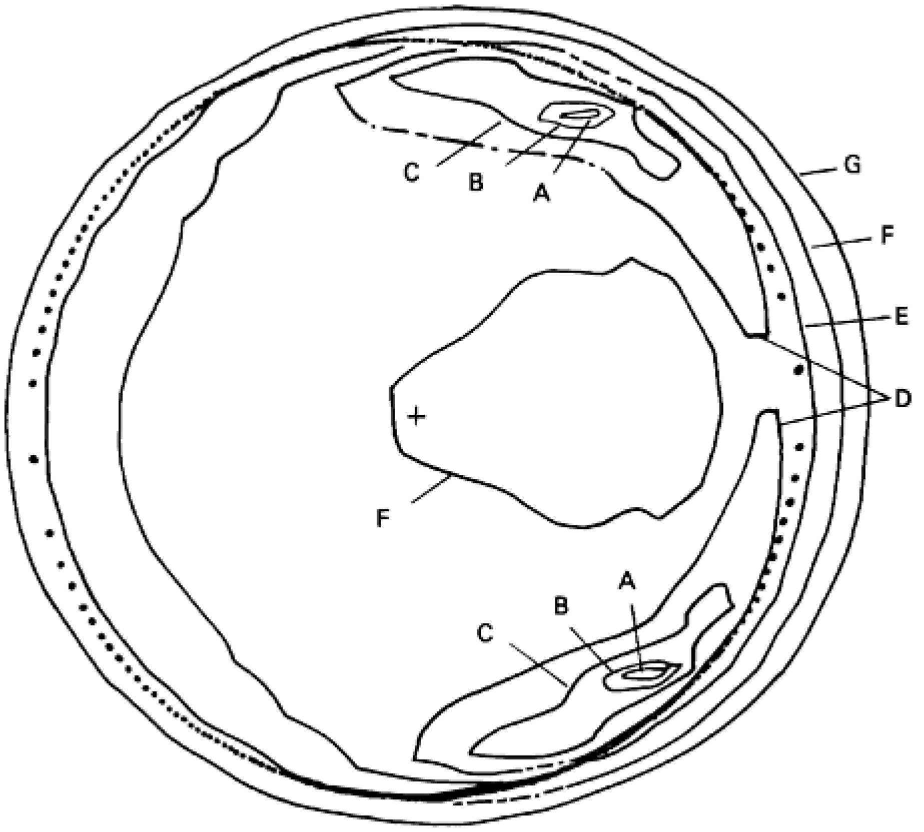


Fig. 8 Contour plot of film thickness, *h/R,* in a point contact. Contour legend: A, 4.0×10^{-6}; B, 4.2×10^{-6}; C, 4.6×10^{-6}; D, 5.2×10^{-6}; E, 6.0×10^{-6}; F, 7.4×10^{-6}; G, 9.0×10^{-6}. Test parameters: *U,* 0.1683×10^{-11}; *W,* 0.1106×10^{-6}; *G* 4.522×10^{3}

More recently, Chittenden *et al.* (Ref 15) extended Hamrock and Dowson's formulas to include the effect of an oblique angle between the entraining velocity and the minor axis of the contact ellipse. They contributed a set of comprehensive formulas for predicting the isothermal minimum and central film thicknesses for more general point contacts. Extensive experimental film measurements have been provided by Koye and Winer (Ref 16) for elliptical conjunctions with the major axis oriented along the rolling direction.

Pressure Distributions. The pressure profile in a line contact deviates slightly from the dry contact elliptical Hertzian profile. It contains a gradual buildup just before the conjunction and a secondary sharp pressure spike just before film termination, as shown in Fig. 9. The height of the pressure spike and its position depend on a nondimensional parameter, U/W^2. For the load level commonly used in hardened gears and rollers, pressure measurements at moderate speeds show no sign of pressure spikes or disturbances at the exit region (Ref 17). Calculated pressure profiles based on full EHL at high loads also show that the distribution is essentially Hertzian, with small inlet and exit disturbances.

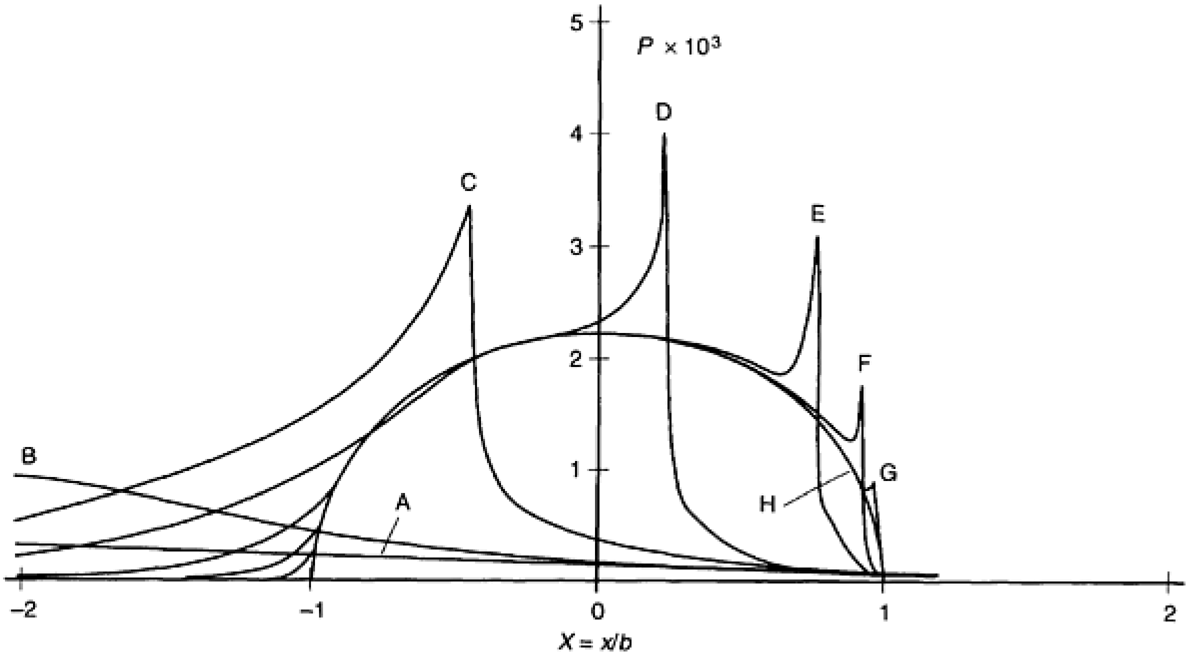


Fig. 9 Pressure distribution for a compressible lubricant at selected surface velocities, *U*. A, *U* = 10^{-8}; B, *U* = $10^{-8.5}$; C, *U* = 10^{-9}; D, *U* = 10^{-10}; E, *U* = 10^{-11}; F, *U* = 10^{-12}; G, *U* = 10^{-13}; H, (dry contact) *U* = 0. Test parameters: W, 3 × 10^{-5}; G, 5.0 × 10^{3}

Sliding Traction and Contact Temperature. In sliding EHL contacts, the tangential force in shearing the lubricant (traction) plays an important role in controlling the skidding of rolling bearings, in influencing fatigue life, and in governing the performance of traction drives. Early traction analyses (Ref 18, 19) and more recent traction models (Ref 20, 21) have revealed that the sliding traction coefficient varies with the sliding speed, as shown in Fig. 10. Traction increases linearly in the low-slip (low slide-to-roll ratio) region, reaches a maximum, and then decreases gradually with the sliding speed. The initial slope and the value of maximum traction depend on rolling speed, viscosity, and load, but the general trend is the same for all conditions. A study of lubricant rheology under high pressure and temperature successfully explained that the lubricant in the conjunction behaves like a viscoelastic substance in the low-slip region and that the viscoplastic behavior can be accurately predicted by a shear stress and shear rate relation containing three constants, all of which can be measured by independent experiments not related to EHL contacts.

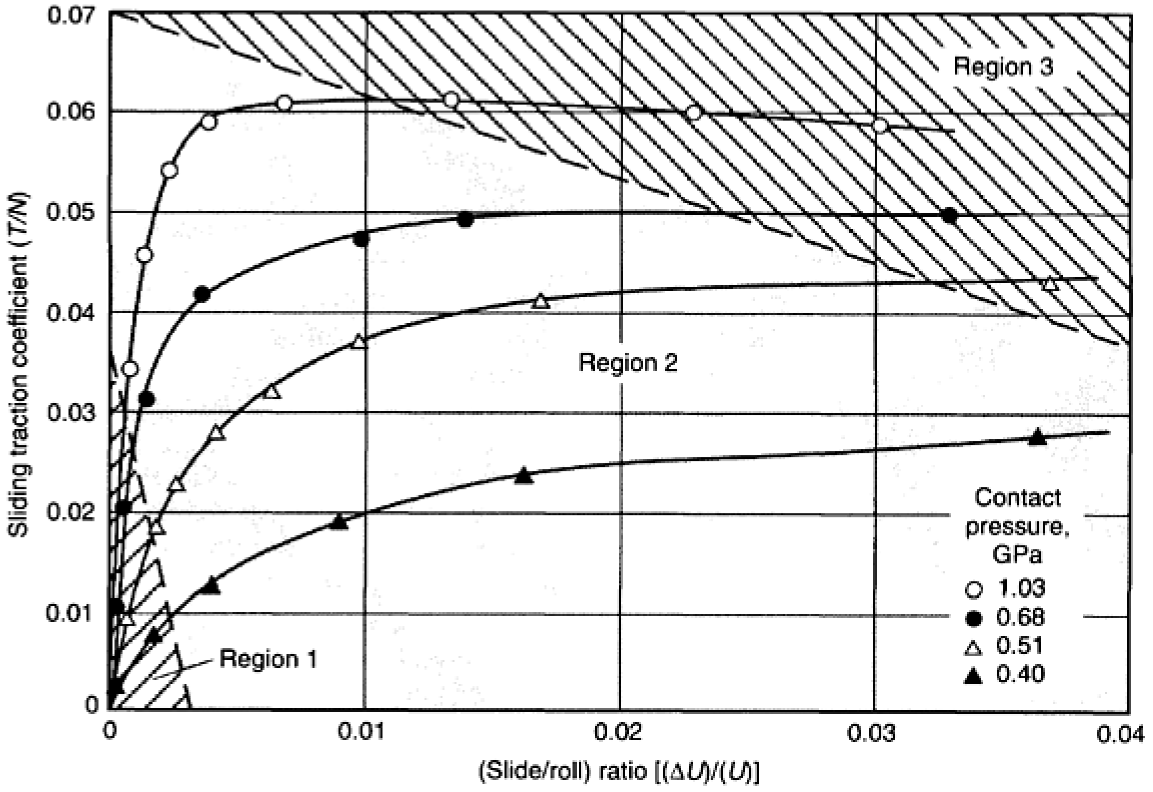


Fig. 10 Typical traction curves obtained at selected mean contact pressures when tested on a two-disk machine. Three distinct traction regions are represented: region 1, linear (that is, low-slip); region 2, nonlinear (isothermal); region 3, thermal. *T,* traction; *N,* normal load; Δ*U,* sliding velocity

The shear stress distribution and sliding traction in EHL contacts can be calculated by Gecim and Winer's model (Ref 22), which relates shear strain rate and shear stress by the equation:

$$\dot{\gamma} = \frac{\dot{\tau}}{G_\infty} + \frac{\tau_L}{\eta}\left[\tanh^{-1}\left(\frac{\tau}{\tau_L}\right)\right] \quad \textbf{(Eq 2)}$$

where the three rheological constants are the limiting shear modulus (G_∞), the limiting shear stress (τ_L), and the static equilibrium viscosity (η), all of which are functions of pressure and temperature and should be measured separately by independent experiments. These relations have been used in practice to determine the local shear stress in line or point contacts. A typical shear stress distribution along the center of an elliptical contact is shown in Fig. 11.

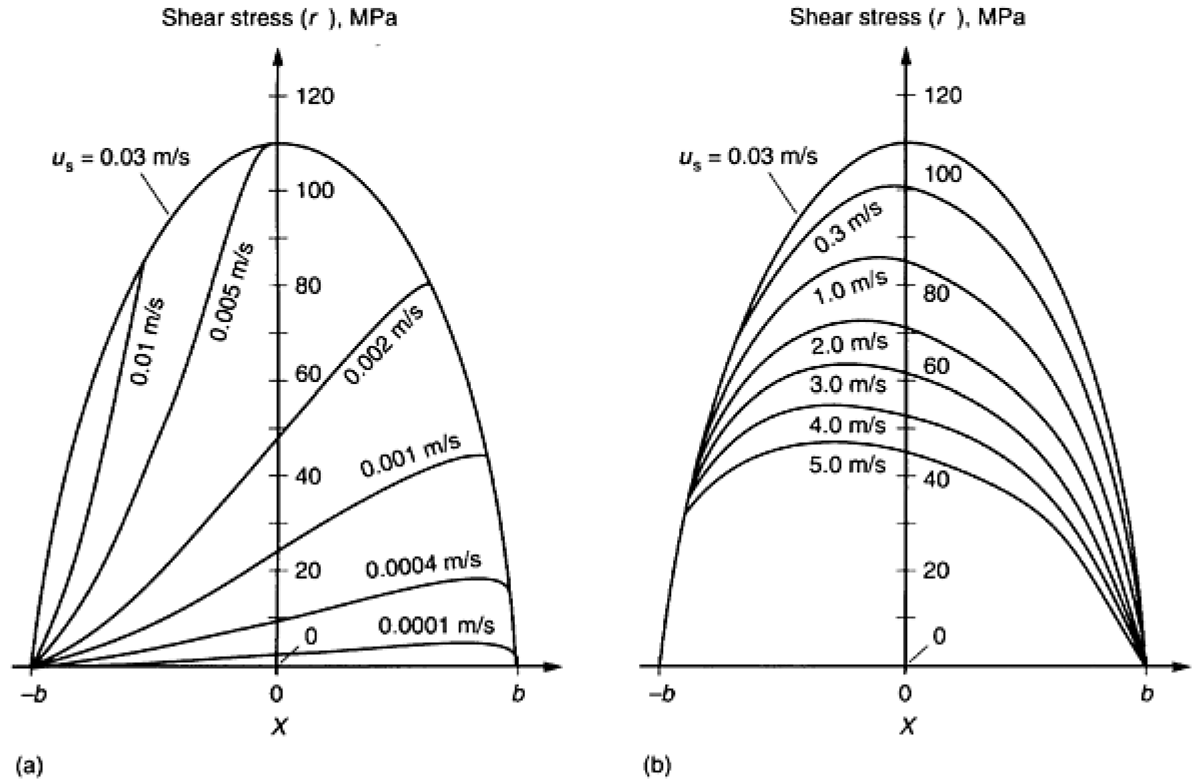


Fig. 11 Typical shear stress distribution along the center of an elliptical contact as a function of sliding condition. (a) Low sliding parameters (1 × 10^{-4} m/s $\leq u_s \leq$ 0.03 m/s). (b) High sliding parameters (0.03 m/s $\leq U_s \leq$ 5.0 m/s). u_s, sliding velocity

Surface temperature in sliding EHL contact is probably the most important variable controlling scuffing and related failures. A surface temperature measurement between a ball and a sapphire plate (Ref 23) confirmed that the Blok-Jaeger method for calculating the maximum surface contact temperature is quite reliable if an accurate sliding coefficient of friction is known. Once the shear stress distribution in full-film EHL has been determined, predicting a thermal map of the surface using Blok-Jaeger's approach is fairly simple. Figure 12 shows typical temperature maps for an elliptical contact with arbitrarily oriented rolling and sliding velocity vectors.

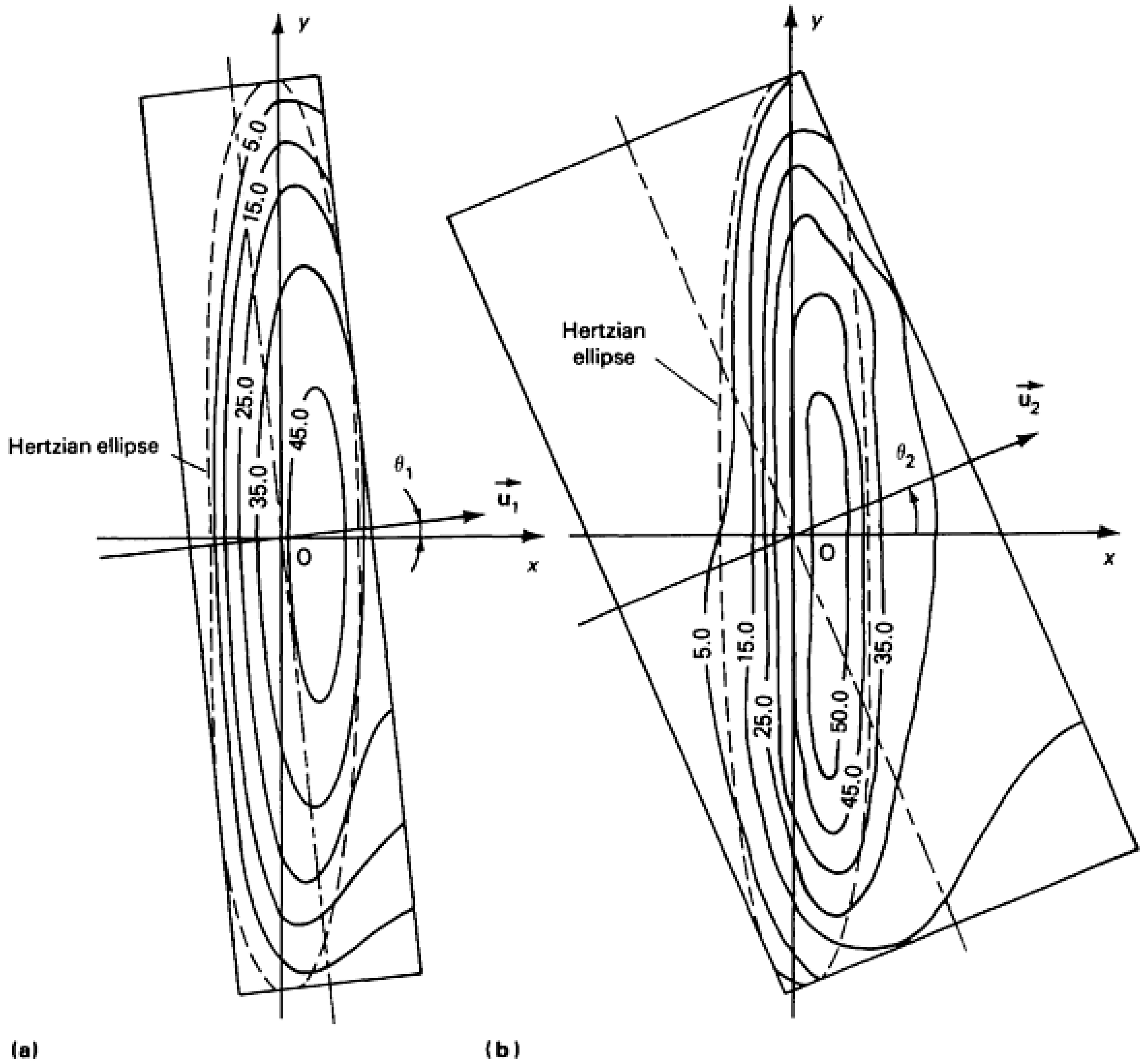


Fig. 12 Typical temperature (T, in °C) maps in an elliptical contact with arbitrarily oriented rolling (θ) and sliding vectors ($\overline{U}$). (a) T_1, θ_1, and U_1. (b) T_2, θ_2, and U_2 (where $T_2 > T_1$; $\theta_2 > \theta_1$; $U_2 > U_1$)

Plastohydrodynamic Lubrication

In metalworking processes, a thick lubricant film can be generated hydrodynamically between the workpiece and the roll or die to control friction and wear. The lubrication behavior between a rigid surface and a plastically deformed surface is known as plastohydrodynamic lubrication. Thick-film PHL occurs infrequently in metalworking lubrication; however, its behavior helps to understand mixed-film PHL, which is more common.

The lubricant film thickness is formed by wedging action just before the workpiece enters the plastically deformed working zone. The ratio of this inlet film thickness to the combined surface roughness of the die and workpiece determines the level of asperity contact at the inlet and throughout the work zone.

For metal rolling and extrusion, the inlet film thickness, h_0, can be calculated from Wilson and Walowit's simple formula (Ref 24), based on a linear wedge profile at the inlet region:

$$h_0 = \frac{6\eta_0\gamma U}{(\tan\theta)(1 - e^{-\gamma\sigma})} \quad \textbf{(Eq 3)}$$

where η_0 is the ambient viscosity of the lubricant, γ is the pressure viscosity coefficient, *U* is the entraining velocity of the surfaces at the inlet zone, θ is the angle of bite, and σ is the workpiece flow strength. The film thickness variation in the work zone depends on the increase in surface velocity caused by the reduction of the workpiece. Additional information is available in the article "Metalworking Lubricants" in this Volume.

Thin-Film Lubrication

Thin-film lubrication of concentrated contacts occupies the regime between thick-film elastohydrodynamic lubrication, where the entire load is supported by the lubricant pressure (regime 3 in Fig. 1), and boundary lubrication, where asperity contacts carry the entire load (regime 2 in Fig. 1). In thin-film lubrication (regime 2 in Fig. 1), the load is shared by the hydrodynamic (or fluid) pressure and the asperity contact pressure. For this reason, thin-film lubrication is sometimes referred to as mixed, partial hydrodynamic, or elastohydrodynamic lubrication.

The boundary between the thick-film and the thin-film regimes can be readily identified by a film parameter, Λ, the ratio of the average lubricant thickness to the composite surface roughness. For surfaces with a Gaussian height distribution, Λ = 3 separates the thick-film and thin-film regimes. The thin-film and boundary lubrication regimes are separated more conveniently by the load ratio, that is, the ratio of the load borne by the fluid pressure to that borne by the asperity contact pressure. The thin-film lubrication regime ends when this load ratio approaches zero. Because of the presence of asperity contacts in thin-film lubrication, the lubrication variables, including film thickness, pressure, shear stress, and surface temperatures, are no longer well-behaved, smooth functions, but instead contain local fluctuations caused by the asperity interactions. For this reason, average quantities--such as the average film, $\bar{h}$--are generally used to describe global or macroscopic variations, and random quantities--such as the asperity film thickness, h^*--are used to describe microscopic variations. Figure 13 illustrates the qualitative features of macrovariables and microvariables in thin-film lubrication.

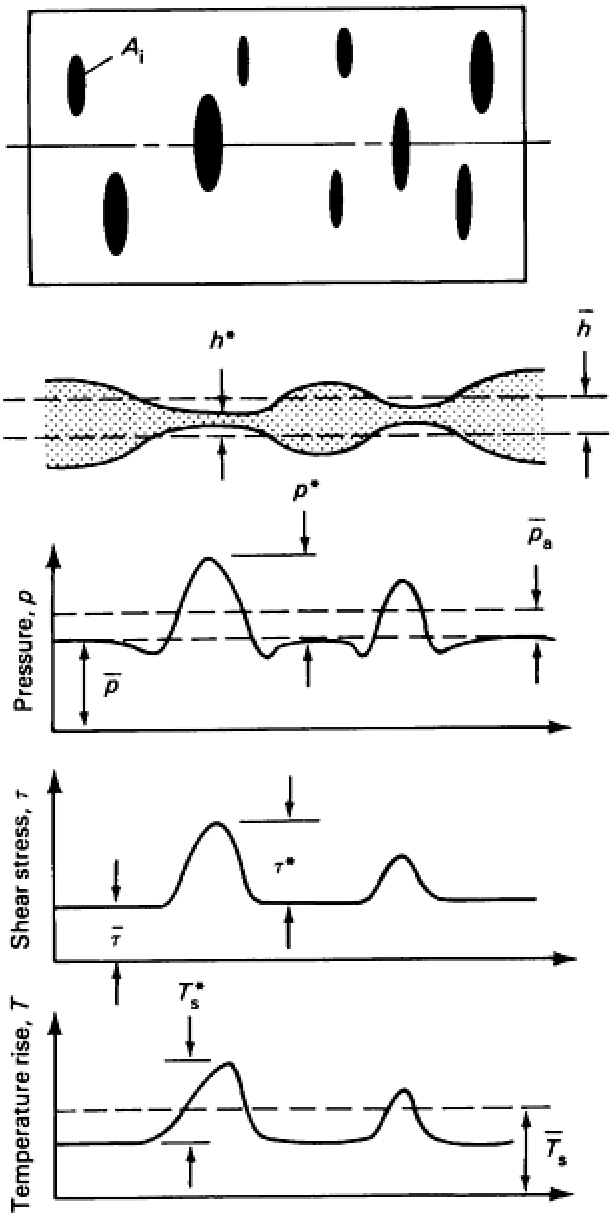


Fig. 13 Qualitative features of macrovariables ($\bar{h}$, $\bar{p}$, $\bar{p}_a$, $\bar{\tau}$, $\bar{T}_s$) and microvariables (h^*, p^*, τ^*, T^*_2) in thin-film lubrication as a function of asperity contacts, A_i

Modes of Asperity Lubrication

The tribological integrity of thin-film lubrication depends on how well the asperities are protected by various modes of lubrication. Figure 14 shows that sliding asperities are protected by three types of lubricating films:

- Micro-EHL and friction polymer films
- Physically adsorbed and other surface films
- Oxide film

Brief descriptions of the generation and breakdown of these films follow.

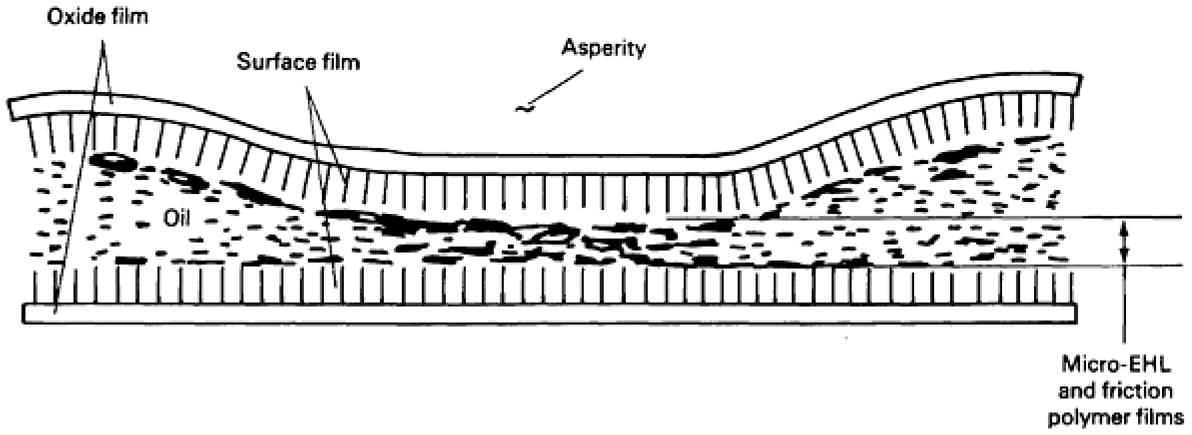


Fig. 14 Three-layer lubricating film that typically protects an asperity from friction and wear

Micro-EHL and Friction Polymer Films. A very thin film of oil can be developed at the asperities of an EHL contact by the normal approaching action of an asperity at the entrance region of the Hertzian conjunction, by the sliding action of the asperity within the conjunction, and by the collision between two sliding asperities. Such films are known as micro-EHL oil films and are the first line of defense against sliding failure. If the geometry of these contacts can be approximated as ellipsoidal tips, these thin oil films can be estimated using existing macro-EHL theories.

The thickness of a micro-EHL film is critically dependent on the lubricant viscosity around the asperity tips. If this viscosity is greatly enhanced by high lubricant pressure generated by the macro-EHL action, then an effective micro-EHL film can be generated: otherwise; the micro-EHL oil film will be too thin. For most mineral oils, the average lubricant pressure, $\overline{p}$, must reach approximately 300 to 500 MPa (45 to 75 ksi) to generate an effective micro-EHL film for protection against scuffing. More detailed descriptions of the breakdown of the macro-EHL action leading to scuffing can be found in Ref 25 and 26.

Another type of polymeric film can also be generated when the asperity temperature becomes excessive because of heating from the sliding friction. This type of film is known as frictionally induced polymeric film. In this process, the lubricant undergoes a primary oxidation and forms a product that, under further oxidation, can polymerize into high-molecular-weight polymers (Ref 27). The polymeric film offers additional protection, but continues to polymerize and finally becomes insoluble sludge and deposit, impeding lubrication.

Physically Adsorbed and Other Surface Films. Physical adsorption is the second line of defense against sliding failure in thin-film lubrication. In this process, a nanolayer or multilayer surface film is formed by adsorption of polar lubricant molecules onto the surface, providing an effective barrier against metal-to-metal contact. This process is reversible. At high temperatures, the adsorbed molecules return to the bulk fluid. When the temperature is lowered, the molecules again become adsorbed on the surface.

The effectiveness of the physically adsorbed surface film ceases when the asperity temperature, T^*_2, reaches a level at which desorption dominates the process. A large fraction of the surface will not be covered by the film, and sliding damage will ensue. For most mineral oils, the critical surface temperature at which half of the surface becomes desorbed is around 150 °C (300 °F). This critical surface temperature appears to depend on the ambient pressure around the sliding asperities (Ref 28). Figure 15 shows a sharp drop in the fractional adsorbate coverage as affected by the ambient lubricant pressure. The critical temperature for $\theta = 0.5$ (that is, half of the surface is uncovered by the surface film) shifts from 140 °C (285 °F) at atmospheric pressure to 250 °C (480 °F) at $\overline{p} = 10^7$ Pa (7×10^4 torr). These data are for poly-α-olefin oils. Similar behavior should also be exhibited by mineral oils. The breakdown of a physically adsorbed surface film generally occurs when the fractional coverage of the adsorbate drops below 0.5. To predict this condition, the asperity temperature, T^*_2, and the ambient lubricant pressure, $\overline{p}$, around the asperity, must be known.

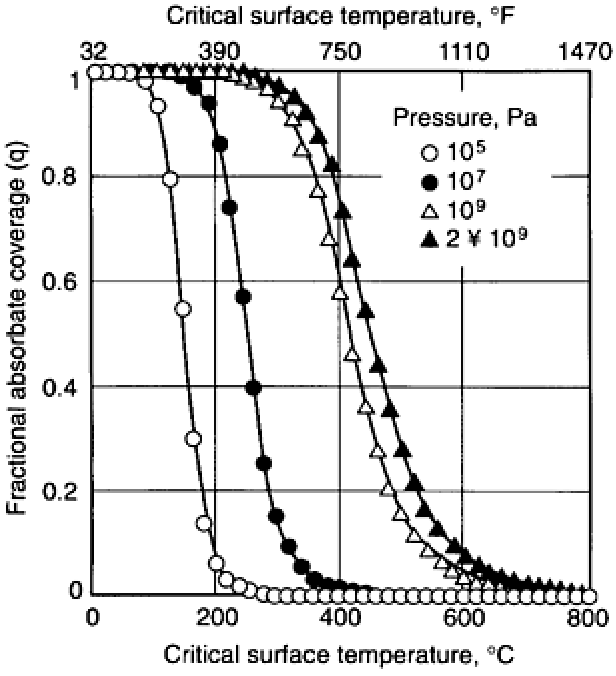


Fig. 15 Plot of fractional adsorbate coverage versus critical surface temperature as a function of lubricant pressure

In addition to adsorbed surface films, other types of surface films can be generated by chemisorption and chemical reaction of the surface with lubricant additives. These films will be discussed in the section "Boundary Lubrication" in this article.

Oxide Film. The last line of defense of the lubricated layers is the oxide film formed by oxidation of the surface with oxygen present in the atmosphere or dissolved in the lubricant. The breakdown mechanisms of oxide films by gradual depletion or by sudden transition from mild to severe wear are only partially understood. The critical conditions for transition from mild to severe wear appear to be associated with a critical speed rather than a critical load (Ref 29).

Boundary Lubrication

Boundary lubrication lies in the regime of extremely low $\Lambda$for cases of very low speed, low viscosity, and very high load. In this regime, the frictional coefficient is usually insensitive to speed, viscosity, or load. The load is totally supported by the asperity contacts. Friction and wear behavior is completely governed by any film that happens to be on the surface, either planned or unplanned. The frictional coefficient, μ, for ferrous surfaces lubricated by nonreactive oils is generally between 0.1 to 0.15. For asperities partially lubricated by surface films, μcan increase to 0.5--the frictional coefficient between most dry sliding iron oxides.

Three types of surface films are commonly known to be effective in boundary lubrication: physically adsorbed, chemically adsorbed, and chemically reacted surface films. Schematic diagrams of the molecular structure of each type of boundary film are shown in Fig. 16 (Ref 30).

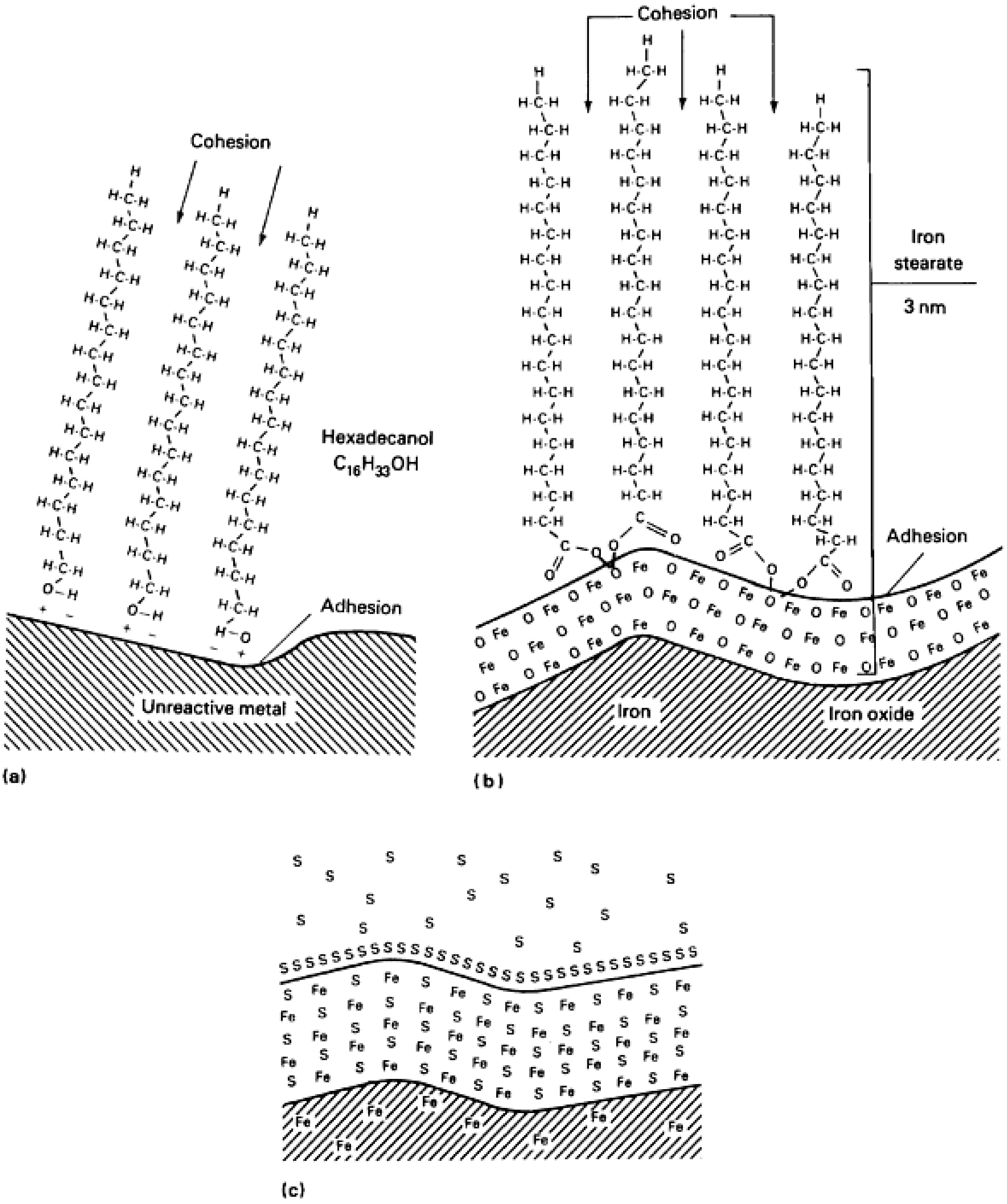


Fig. 16 Schematics showing three types of surface films present in boundary lubrication. (a) Physisorption with preferred orientation of three polar molecules of hexadecanol to a metal surface. (b) Chemisorption of stearic acid on an iron surface to form a monolayer of iron stearate, a soap. (c) Inorganic film formed by chemical reaction of sulfur with iron to form iron sulfide. Source: Ref 30

Adsorbed films are usually monomolecular and have thicknesses in the range of nanometers. Reactive films can be quite thick and are much more stable and durable than adsorbed films. They are enhanced by the additives in the base lubricants that form sulfides, chlorides, or phosphates with the solid surfaces, providing effective surface films for protection of sliding asperities under extreme conditions.

References

1. H.G. Elrod, A Review of Theories for the Fluid Dynamic Effects of Roughness on Laminar Lubricating Films, *4th Leeds-Lyon Symposium on Tribology,* Elsevier, 1978
2. H.S. Cheng, Surface Roughness Effects in Lubrication, *11th Leeds-Lyon Symposium on Tribology,* Elsevier, 1984
3. N. Patir and H.S. Cheng, Effect of Surface Roughness Orientation on the Central Film Thickness in EHD Contacts, *5th Leeds-Lyon Symposium on Tribology,* Elsevier, 1979
4. J.A. Greenwood and J.H. Tripp, The Contact of Two Nominally Flat Rough Surfaces, *Proc. Inst. Mech. Eng.,* Vol 185, 1970-1975, p 625-633
5. A.A. Raimondi *et al.,* Analysis and Design of Sliding Bearings, *Standard Handbook of Lubrication Engineering,* McGraw-Hill, 1968, Chapter 5
6. R.D. Arnell *et al., Tribology, Principles and Design Applications,* Springer-Verlag, 1991, Chapter 5, 6
7. Z. Satar and A.Z. Szeri, Thermal Hydrodynamic Lubrication in Laminar and Turbulent Regimes, *Trans. ASME (Series F),* Vol 96, 1974, p 48-56
8. D. Wymer and A. Cameron, EHD Lubrication of a Line Contact, *Proc. Inst. Mech. Eng.,* Vol 188, 1974, p 221
9. D. Dowson, Elastohydrodynamic Lubrication, *Interdisciplinary Approach to the Lubrication of Concentrated Contacts,* NASA Special Publication 237, 1974, p 34
10. L.E. Murch and W.R.D. Wilson, A Thermal Elastohydrodynamic Inlet Zone Analysis, *J. Lubr. Technol. (Trans. ASME),* Vol 97, 1975, p 212
11. B.J. Hamrock and D. Dowson, Isothermal Elastohydrodynamic Lubrication of Point Contacts. Part I--Theoretical Formulation, *J. Lubr. Technol. (Trans. ASME),* Vol 98, 1976, p 223
12. B.J. Hamrock and D. Dowson, Isothermal Elastohydrodynamic Lubrication of Point Contacts. Part II--Ellipticity Parameter Results, *J. Lubr. Technol. (Trans. ASME),* Vol 98, 1976, p 245
13. B.J. Hamrock and D. Dowson, Isothermal Elastohydrodynamic Lubrication of Point Contacts. Part III--Fully Flooded Results, *J. Lubr. Technol. (Trans. ASME),* Vol 99, 1977, p 264
14. B.J. Hamrock and D. Dowson, Isothermal Elastohydrodynamic Lubrication of Point Contacts. Part IV--Starvation Results, *J. Lubr. Technol. (Trans. ASME),* Vol 99, 1977, p 15
15. R.J. Chittenden, D. Dowson, J.F. Dunn, and C.M. Taylor, A Theoretical Analysis of the Isothermal Elastohydrodynamic Lubrication of Concentrated Contacts. II. General Case, With Lubricant Entrainment Along Either Principal Axis of the Hertzian Contact Ellipse or at Some Intermediate Angle, *Proc. R. Soc. (London) A,* Vol 397, 1985, p 271-294
16. K.A. Koye and W.O. Winer, "An Experimental Evaluation of the Hamrock and Dowson Minimum Film Thickness Equation for Fully Flooded EHL Point Contacts," Paper 80-C2-Lub-38, American Society of Mechanical Engineers
17. J.W. Kannel *et al.,* "A Study of the Influence of Lubricants on High-Speed Rolling-Contact Bearing Performance," Technical Report ASD-TR-61-643, Part VI, 1966
18. K.L. Johnson and R. Cameron, Shear Behavior of EHD Oil Film at High Rolling Contact Pressures, *Pro. Inst. Mech. Eng.,* Vol 182, 1967, p 307
19. A. Dyson, Frictional Traction and Lubricant Rheology in Elastohydrodynamic Lubrication, *Philos. Trans. R. Soc. (London),* Vol 266, 1970, p 1170
20. K.L. Johnson and J.L. Tevaarwerk, Shear Behavior of EHD Oil Films, *Proc. R. Soc. (London) A,* Vol 356, 1977, p 215
21. S. Bair and W.O. Winer, Shear Strength Measurements of Lubricants at High Pressure, *J. Lubr. Technol. (Trans. ASME),* Vol 101, 1979, p 251
22. B. Gecim and W.O. Winer, A. Rheological Model for Elastohydrodynamic Contacts Based on Primary Laboratory Data, *J. Lubr. Technol. (Trans. ASME),* Vol 101, 1979, p 258-265
23. H.S. Nagaraj, D.M. Sanborn, and W.O. Winer, Direct Surface Temperature Measurements by Infrared

Radiation in EHD Contacts and the Correlation With the Block Temperature Theory, *Wear,* Vol 49, 1978, p 43

24. W.R.D. Wilson and J. Walowit, An Isothermal Hydrodynamic Lubrication Theory for Hydrostatic Extrusion and Drawing Processes With Conical Dies, *J. Lubr. Technol. (Trans. ASME),* Vol 93, 1971, p 69-74
25. A. Dyson, The Failure of Elastohydrodynamic Lubrication of Circumferentially Ground Discs, *Proc. Inst. Mech. Eng.,* Vol 190 (No. 1), 1976, p 52-76
26. H.S. Cheng and A. Dyson, Elastohydrodynamic Lubrication of Circumferentially Ground Discs, *ASLE Trans.,* Vol 21 (No. 1), 1978, p 25-40
27. D.B. Clark, E.E. Klaus, and S.M. Hsu, The Role of Iron and Copper in the Oxidation Degradation of Lubricating Oils, *Lubr. Eng.,* Vol 41 (No. 5), 1985, p 280-289
28. S.C. Lee and H.S. Cheng, Correlation of Scuffing Experiments with EHL Analysis of Rough Surfaces, *J. Tribology (Trans. ASME),* accepted for publication
29. S.C. Lim and M.F. Ashby, *Acta Metall.,* Vol 35, 1987, p 1
30. D. Godfrey, Boundary Lubrication, *Interdisciplinary Approach to Friction and Wear,* P.M. Ku, Ed., NASA Special Publication SP-181, 1968, p 335-353

Lubricant Additives and Their Functions

Syed Q.A. Rizvi, The Lubrizol Corporation

Introduction

A LUBRICANT protects and prolongs the life of equipment by performing these important functions (Ref 1):

- Lubrication, because it reduces friction and wear by introducing a lubricating film between moving parts
- Cooling, because it helps dissipate heat away from the critical parts of the equipment
- Cleaning and suspending, because it facilitates the smooth operation of the equipment by removing and suspending deposits such as carbon, sludge, and varnish
- Protection, because it prevents metal damage that is due to oxidation and corrosion

Untreated, or nonformulated, lubricants (mineral-based oils and synthetic-based oils) do not possess the necessary properties to be effective in the demanding lubrication environments that exist today. To perform the above-mentioned functions properly, base fluids need the help of chemical additives. Additives improve the lubricating ability of the base oils by either enhancing the desirable properties already present or adding new properties. For this reason, additives are an integral part of modern formulated lubricants.

A formulated lubricant comprises a base fluid (Ref 2) and a performance package. The performance package contains a number of additives that help improve the lubricating ability of the base fluid. The quality and quantity of additives in the performance package depend on the quality of the base fluid and the intended use. In general, poor-quality base fluids need better additives, and possibly in larger amounts, than the base fluids of good quality. Likewise, applications that put more demand on the lubricant (engine oils and automotive gear oils) require superior additives than less-demanding applications, such as industrial lubricants and metalworking fluids.

Lubricants can be broadly classified as either engine or nonengine lubricants. Engine lubricants can be subcategorized, in terms of application, as:

- Gasoline engine oils

- Diesel engine oils (automotive, stationary, railroad, and marine)
- Natural gas engine oils
- Aviation engine oils
- Two-stroke cycle engine oils
- Alternate fuel engine oils

Nonengine lubricants are subcategorized as:

- Transmission fluids (automatic, manual, and power)
- Power steering fluids
- Shock absorber fluids
- Gear oils (automotive and industrial)
- Hydraulic fluids (tractor and industrial)
- Metalworking fluids
- Miscellaneous industrial oils
- Greases

Engine lubricants or oils are designed for use in internal combustion engines that operate on a wide variety of fuels. Major problems in these engines relate to the interaction of combustion products with the lubricant, which results in deposits and wear. The combustion products find their way into the crankcase as "blowby" and contaminate the lubricant. This causes the formation of a wide variety of oxygenated products through lubricant autoxidation. The interaction between the components of the blowby and these oxygenated products leads to the formation of harmful engine deposits which, if left unchecked, ultimately impair the operation of an engine (Ref 3, 4).

The performance of engine oils is judged on their ability to reduce friction, resist oxidation, minimize deposit formation, and prevent corrosion and wear. In the United States, the performance specifications for engine oils are established by the collaborative efforts of the Society of Automotive Engineers (SAE), the American Petroleum Institute (API), and the American Society for Testing and Materials (ASTM) (Ref 5). The API establishes the performance categories or service classifications and their descriptions for engine oils. The U.S. military and original equipment manufacturers (OEMs) have their own performance requirements, which can be over and above those of the API. Other countries have similar organizations that establish performance criteria and categories for lubricants used within their domains.

The SAE viscosity grades and API service classifications are the usual means used to define the performance of automotive engine lubricants (Ref 6, 7). The viscosity classification assures easy starting and lubrication at low temperatures, and adequate lubrication at high temperatures. The service classification assures the user that the lubricant meets the performance requirements of the equipment manufacturer.

Nonengine lubricants are used to lubricate equipment that operates in a noncombustion environment. Transmission fluids, hydraulic fluids, gear oils, and greases are used for mechanisms that transfer power from a power source, such as an engine, to parts that perform the actual work, such as wheels.

Transmission fluid performance requirements fluids are established by OEMs (Ref 8). The most important features of the transmission fluids are their frictional consistency (durability) and frictional compatibility with the transmission components. In automatic transmissions, such components include clutches and bands; manual transmissions and manual transaxles also include cone- or plate-type synchronizers. Unlike automatic transmissions, which use transmission fluids recommended by OEMs only, manual transmissions use a wide variety of fluids, including automatic transmission fluids, engine oils (5W-30), some gear lubricants, and specialty fluids. In addition to frictional properties, certain OEMs require that the transmission fluids used in their equipment must have improved shear stability, low-temperature fluidity, and other specific characteristics.

Power transmission fluids are designed for use in heavy-duty automatic transmissions and torque converters, which are commonly used in agriculture, transportation, and construction equipment. The viscosity and frictional properties of these fluids are critical to their performance. Both SAE and OEM performance specifications are used to describe these fluids (Ref 9).

Power steering fluid performance requirements are established by the OEMs. The quality is assessed on the basis of frictional properties, seal compatibility, oxidation resistance, and rust control. Some automatic transmission fluids closely meet these criteria and can be used in this application.

Shock absorber fluids require good frictional characteristics, oxidation resistance, wear protection, and seal compatibility. Compressibility and antifoaming characteristics are also desirable. The shock absorber producer and the OEMs jointly determine the performance criteria for these fluids.

Gear oils are formulated to provide both gears and axles with extreme-pressure protection against fatigue, scoring, and wear under boundary lubrication conditions (Ref 10).

Automotive gear oils are classified in a manner similar to engine oils; that is, through SAE viscosity grades (Ref 11), API service designations (Ref 12, 13), U.S. military specifications, and OEM performance requirements. However, the gear oil viscosity grading system is different from the engine oil viscosity grading system.

The service requirements of industrial gear oils are established by the American Gear Manufacturers Association (AGMA), the Society of Tribologists and Lubrication Engineers (STLE), and a variety of other organizations, such as U.S. Steel, Cincinnati Milacron, and Alcoa (Ref 14).

Hydraulic fluids are of two general types: those used to lubricate tractor hydraulics and those used to lubricate industrial hydraulic equipment.

Tractor hydraulic fluids combine hydraulic and transmission properties with extreme-pressure properties. Tractor hydraulic fluids differ widely in performance requirements because OEMs cannot agree on common specifications for a universal tractor hydraulic fluid. These fluids, in general, seek to provide extreme-pressure and wear protection, and to match the frictional requirements of the equipment. The quality of these fluids is assessed on the basis of their ability to meet individual OEM specifications, as well as these performance requirements: API GL-4 (for extreme pressure), Caterpillar TO-2/TO-4 (for friction), and Allison C-4 (for friction, oxidation, and wear).

Industrial hydraulic fluids comprise three categories: antiwear hydraulic fluids, fire-resistant fluids, and rust and oxidation-inhibited oils (R&O oils). The OEMs define the performance criteria for these lubricants. Each hydraulic pump manufacturer has its own performance requirements governing viscosity, antiwear properties, and ability to inhibit rust, oxidation, corrosion, foam, filter plugging, and demulsibility.

Metalworking fluids are used to facilitate the manufacture of metal parts. They do so by providing lubrication during workpiece formation, removing debris and heat, and aiding in the cutting, grinding, and shaping operations. Based on their functions, these fluids are classified as forming fluids, removal fluids, protecting fluids, and treating fluids. The specifications for these fluids are established by OEMs and the end users.

Miscellaneous industrial oils include turbine oils, slideway lubricants, rock drill oils, and the like. These lubricants are usually specified by the International Organization for Standardization (ISO) viscosity grades and U.S. military and (OEM) performance requirements.

Greases are classified by chemical composition and many other properties. They are commonly identified by the type of thickener used, by the National Lubricating Grease Institute (NLGI) consistency grades, and by NLGI service classifications (Ref 15).

Additives

Lubricant additives can be broadly categorized as being either chemically active or chemically inert. Chemically active additives, such as dispersants, detergents, antiwear and extreme-pressure (EP) agents, oxidation inhibitors, and rust and corrosion inhibitors, have the ability to interact chemically with metals (to form a protective film) and the polar oxidation and degradation products (to make them innocuous). Chemically inert additives, which improve the physical properties that are critical to the effective performance of the lubricant, include emulsifiers, demulsifiers, pour-point depressants, foam inhibitors, and viscosity improvers.

Most lubricant additives, except perhaps some viscosity improvers and pour-point depressants, consist of an oleophilic hydrocarbon group and a hetero atom (N,O,S, and P)-based polar functionality (Fig. 1). The hydrocarbon group is of

sufficient carbon length to impart the desired solubility characteristics to the additive. The additives that require greater solubility in oil (dispersants, detergents, and viscosity improvers) usually contain large hydrocarbon groups. Those that require either lower solubility or greater surface activity (foam inhibitors and EP agents) contain small hydrocarbon groups. A proper balance of polar and nonpolar characteristics is critical to the additive's performance.

Fig. 1 Typical additive molecule

The functional groups of chemically active and inert additives identified above, as well as friction modifiers and other additives, are described more fully below.

Dispersants

Dispersants are additives that are used to suspend oil-insoluble resinous oxidation products and particulate contaminants in the bulk oil. By doing so, they minimize sludge formation, particulate-related abrasive wear, viscosity increase, and oxidation-related deposit formation.

A dispersant molecule contains an oleophilic hydrocarbon moiety and a polar functional group. The polar group, usually oxygen- or nitrogen-based, attaches itself to the oxidation products and sludge particles, while the oleophilic hydrocarbon group keeps the particles suspended in oil (Ref 4). The hydrocarbon radical is either oligomeric or polymeric, and is usually aliphatic in nature. It contains from 70 to 200 or more carbon atoms to ensure good oil solubility, with a polybutenyl alkyl group being the most commonly used hydrocarbon group.

In some dispersants, the hydrocarbon moiety is derived from a high molecular weight polymer, such as olefin copolymer, polyacrylate, or styrene-ester polymer. Such dispersants can function as dispersants as well as viscosity improvers, and are appropriately called dispersant-viscosity improvers.

The chemical classes that are suitable for use as dispersants include alkenylsuccinimides, succinate esters, high molecular weight amines, Mannich bases, and phosphonic acid derivatives. Commercially, polybutenylsuccinic acid derivatives are the most commonly used dispersant types.

Both the succinimides and the succinate esters are prepared from polybutenylsuccinic anhydrides. Polyisobutylene (molecular weight between 440 and 5000) is reacted with maleic anhydride, either thermally or in the presence of chlorine, to yield the succinic anhydride. Further reaction of the succinic anhydride with amines (alkylenepolyamines and heterocyclic polyamines) results in the formation of succinimide dispersants. The reaction of the succinic anhydride with alcohols (especially polyhydric alcohols) results in the formation of succinate-ester dispersants. Figure 2 summarizes the syntheses of these dispersants.

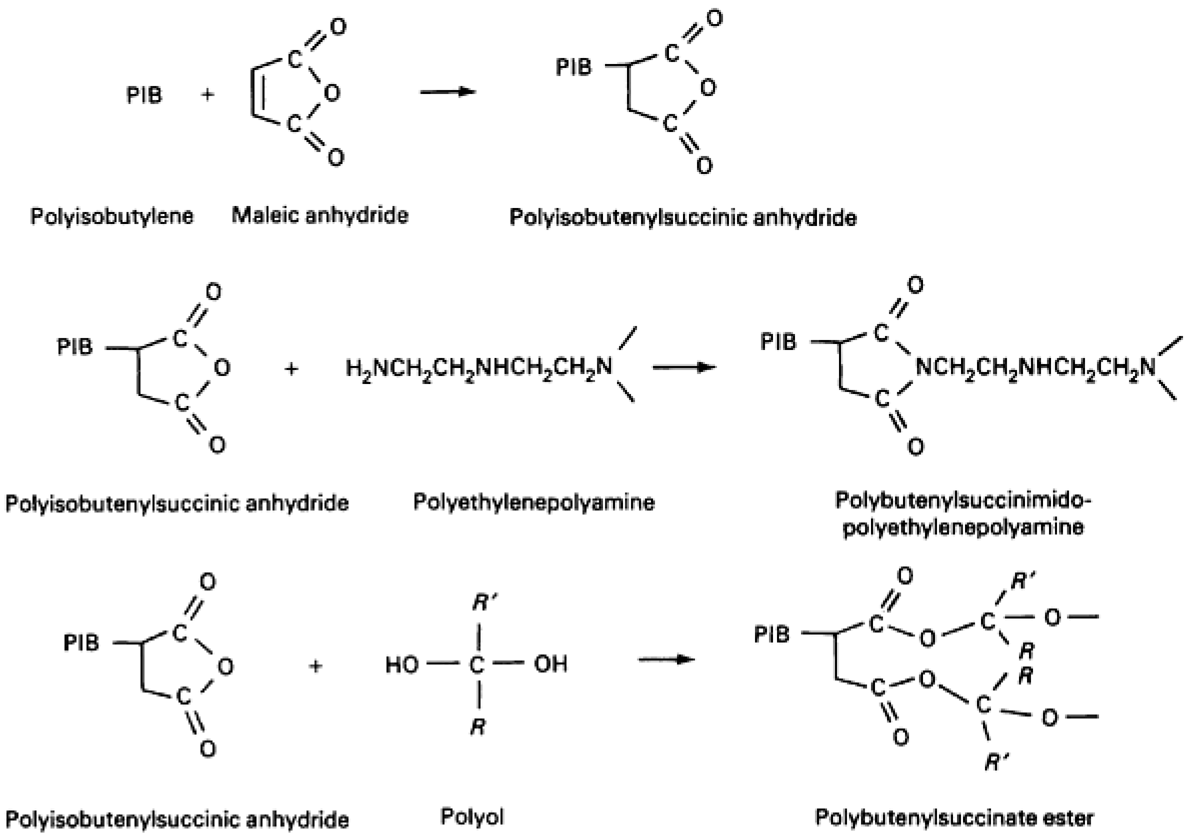


Fig. 2 Synthesis of alkenylsuccinic acid dispersants

Mannich dispersants are produced by the condensation of a high molecular weight alkylphenol (polybutenylphenol), an aldehyde, and alkyleneamines or alkylenepolyamines. Phosphonic acid ester dispersants are prepared by reacting phosphonic or thiophosphonic acids with either ethylene oxide or propylene oxide. The starting acids are obtained from the hydrolysis of the olefin-phosphorus pentasulfide adducts. The preparation of the Mannich products and the phosphonic esters is shown in Fig. 3.

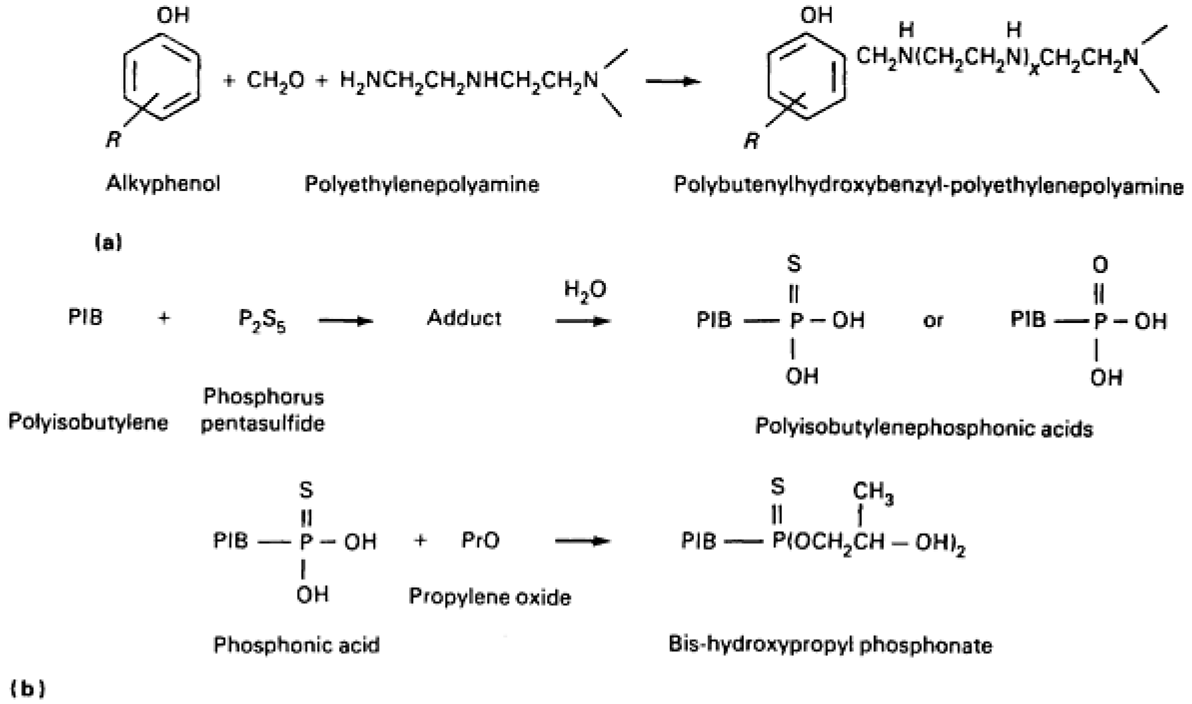


Fig. 3 Synthesis of dispersants (a) Mannich. (b) Phosphonic acid

The dispersants are primarily used in gasoline engine oils, diesel (heavy-duty and railroad) engine oils, natural gas engine oils, and aviation piston engine oils. Dispersants are also used in automatic transmission fluids and gear lubricants. The gasoline and heavy-duty diesel engine oils account for 75 to 80% of the total dispersant use.

Succinimide and succinate-ester (pentaerythritol esters of polyisobutenylsuccinic anhydride) types are used both in gasoline and diesel engine oils. High molecular weight amine and Mannich types are used in gasoline engine oils only. Succinimide dispersants also find use in automatic transmission fluids, power steering fluids, and, on a limited basis, in gear oils.

In gasoline and diesel engine oils, the effectiveness of a dispersant is assessed on the basis of its ability to disperse lamp black or used engine oil sludge in laboratory screen tests and its performance in various ASTM sequence dynamometer engine tests, as well as fleet testing.

Detergents

Detergents perform functions that are similar to those of dispersants. Additionally, detergents neutralize acidic combustion and oxidation products, and hence, control rust, corrosion, and resinous build-up in the engine.

Detergents are metal salts of organic acids. The quantity of metal can either equal or exceed the stoichiometric amount necessary for complete neutralization of the acid function. When the metal is present in a stoichiometric amount, the detergents are referred to as neutral; when excess metal is present, they are referred to as either basic or overbased.

Detergents are described chemically in terms of their soap content, the degree of overbasing, and the total base number (TBN), which is expressed as mg KOH/g of additive. The soap content refers to the amount of neutral salt and reflects cleaning ability, or detergency. The degree of overbasing describes the ratio of equivalents of metallic base to equivalents of acid substrate. The TBN of the detergent indicates its acid neutralizing ability.

Detergents, like dispersants, contain a surface-active polar group and an oleophilic hydrocarbon radical with an appropriate number of carbon atoms to ensure good oil solubility. Metal sulfonates, phenates, carboxylates, salicylates, and phosphonates are common examples of the polar groups that are present in detergent molecules.

Alkylbenzenesulfonic acids, alkylphenols, sulfur-coupled and methylene-coupled alkylphenols, carboxylic acids, and alkylphosphonic acids are the commonly used detergent substrates. The methods of their synthesis are shown in Fig. 4. Alkylbenzenesulfonic acids are obtained by sulfonating alkylbenzenes. The products obtained by sulfonating alkylaromatics from petroleum refining are referred to as natural sulfonates, and those obtained by sulfonating alkylaromatics from catalytic alkylation process are referred to as synthetic sulfonates. Alkylphenols are prepared from phenol and an olefin by using an acid catalyst. These alkylphenols can be further reacted with sulfur, sulfur dichloride, or formaldehyde to form sulfur-bridged and methylene-bridged alkylphenols. Alkylsalicylic acids are prepared through the use of Kolbe's process, which involves reacting an alkali metal phenate with carbon dioxide. Alkylphosphonic acids are the hydrolysis products of polyisobutylene-phosphorus pentasulfide adducts.

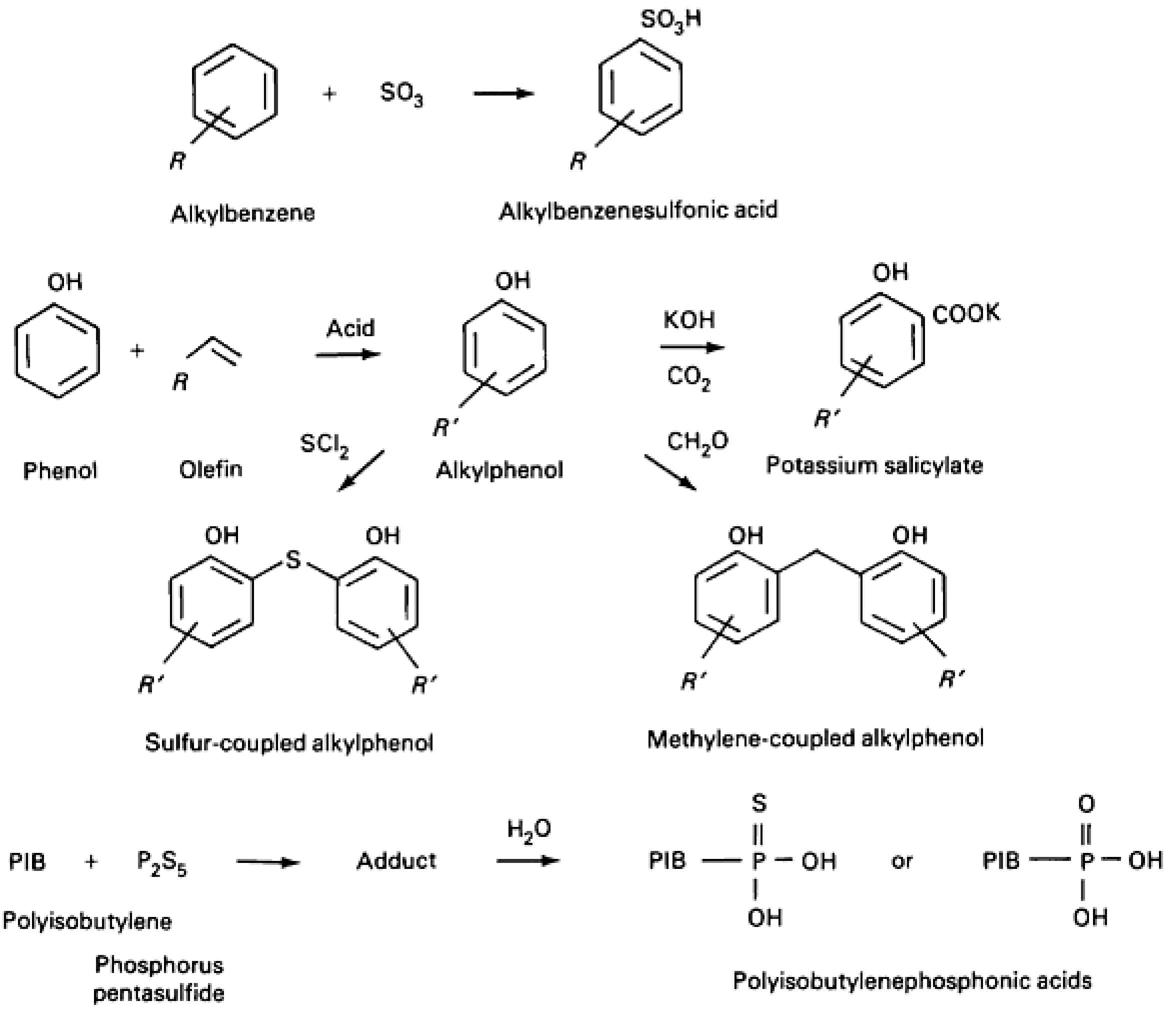


Fig. 4 Detergent substrates

The detergents are prepared by reacting an organic acid with an appropriate metal base in the presence of a polar promoter. A number of metals can be used to make detergents, but sodium, magnesium, calcium, and barium are the most common. Of these, calcium and magnesium are metals that are used most widely, with a clear preference for calcium because of its lower cost. Use of barium is being phased out because of toxicity concerns. The bases of choice are caustic (sodium hydroxide) for sodium detergents; lime (calcium hydroxide) for calcium detergents; magnesium oxide for magnesium detergents; and barium hydroxide for barium detergents. In basic detergents, which contain excess base, the base can be present as is or as metal carbonate. In practice, virtually all commercially available detergents are overbased to some extent. For example, commercial "neutral" sulfonates have a TBN of 30 or less. "Basic" detergents typically have a TBN that ranges from 200 to 500.

Sulfonate, salicylate, and carboxylate detergents are commercially available as calcium and magnesium salts. Phenate detergents are available as calcium salts, and phosphonate detergents are available as barium salts only. Basic calcium sulfonates constitute 65% of the total detergent market, followed by phenates at 31%.

The preparation of various detergent types is depicted in Fig. 5, whereas Fig. 6 shows the generalized structures for the neutral detergents. Basic detergents can be considered neutral detergents that contain the excess base in an associated form.

Substrate + Metal Oxide or Hydroxide (MO, $M(OH)_n$; Stoichiometric amount) $\longrightarrow$ Neutral Salt

Substrate + Metal Oxide or Hydroxide (MO, $M(OH)_n$; Excess) $\xrightarrow[\text{or no } CO_2]{CO_2}$ Basic or Overbased Salt

Basic Salts = Neutral Salts . MO, $M(OH)_n$ or $M_n CO_3$

M = Na, Mg, Ca, Ba   n = 1 or 2

Fig. 5 Preparation of detergents

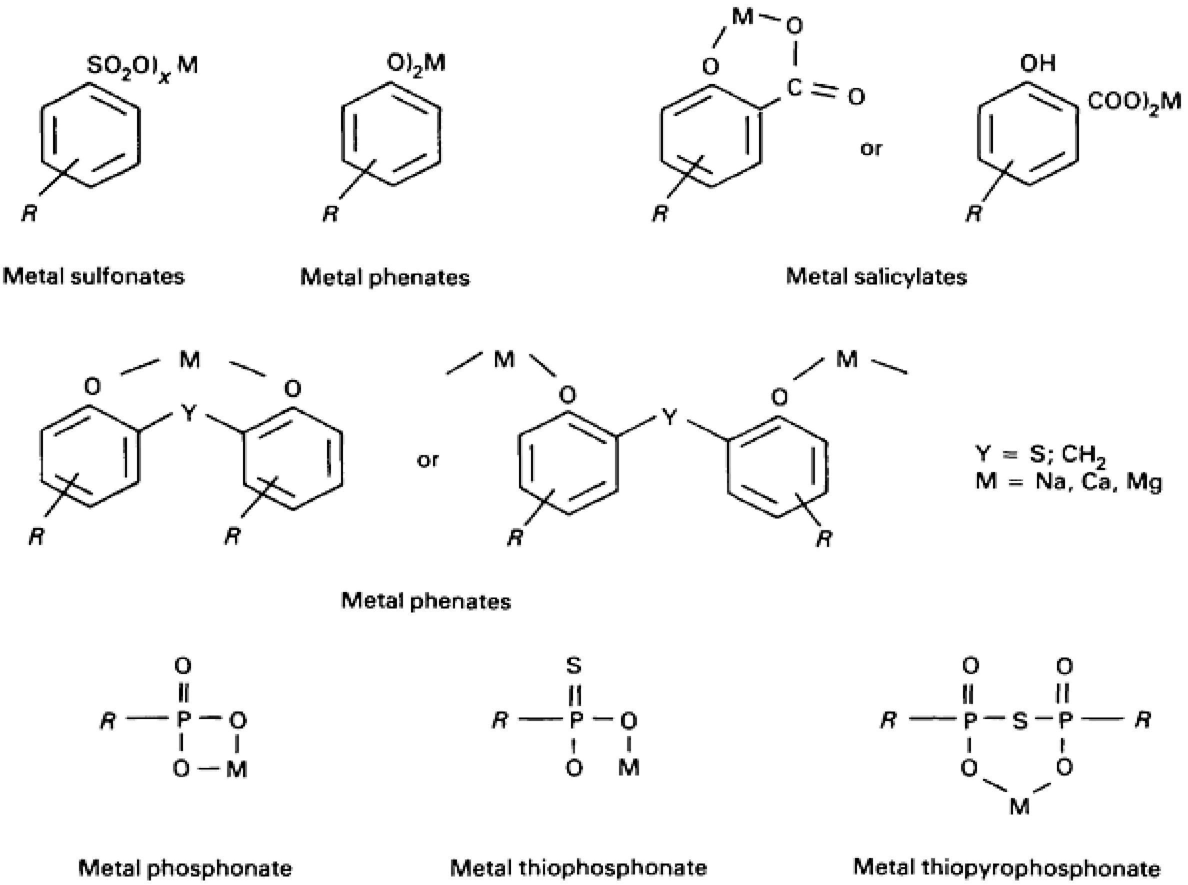


Fig. 6 Idealized structures for neutral detergents

Detergents are used in lubricants to keep oil-insoluble by-products of combustion and oil oxidation in suspension. Depending on the end use of the lubricant, insoluble by-products can be coke, acidic decomposition products, or deposit-forming resinous products (Ref 3, 4). Detergents control the buildup of these undesirable contaminants by keeping the equipment surfaces clean and by neutralizing the acidic products of lubricant oxidation and decomposition. They keep the equipment surfaces clean, both by forming a protective film on the metal surfaces through adsorption and by suspending the deposit-forming species in the bulk oil through association (Ref 16).

Detergents can act as oxidation inhibitors, depending on the nature of the functional group. For example, phenates, sulfurized phenates, and salicylates possess oxidation-inhibiting properties, presumably because of the presence of the phenolic functional group.

Detergents are primary used in crankcase lubricants. Gasoline and diesel engine oils account for over 75% of the total detergent consumption. Detergent treatment levels in engine lubricant formulations are fairly high; marine diesel engine lubricants contain the highest detergent concentrations. Marine engines use high-sulfur fuel, which leads to acidic combustion products (sulfuric acid) and, therefore, need the high base reserve in "basic" detergents for neutralization.

Detergency in diesel engine oils (for API quality) is evaluated using both single-cylinder engine tests (Caterpillar 1K/1H2) and multicylinder engine tests (NTC-400, Mack T-6). Some OEMs use their own specifications. Detergents have additional use in automatic transmission fluids and tractor hydraulic fluids. However, their use in these applications is not to control deposits, but to modify the frictional properties of the fluid.

Antiwear and Extreme-Pressure Agents

Wear is a phenomenon that occurs in all equipment that has moving parts. Three conditions that can lead to wear are surface-to-surface contact (frictional wear); surface contact with foreign matter (abrasive wear); and erosion by corrosive materials (corrosive wear). Abrasive wear can be prevented by installing an efficient filtration mechanism to remove the offending debris. Corrosive wear can be controlled by using additives that can neutralize the reactive species that attack the surface.

Under normal conditions of speed and load, two metal surfaces are effectively separated by a lubricant film, a condition known as hydrodynamic, or thick-film, lubrication. An increase in load or a decrease in speed promotes metal-to-metal contact. This causes a rise in temperature in the contact zone, which is due to frictional heat. The consequence is lubricant viscosity loss, which decreases the film-forming ability of the lubricant and the ability to minimize metal-to-metal contact. Under these conditions, the nature of lubrication changes from hydrodynamic to mixed-film to boundary, or thin-film, lubrication. Antiwear additives and EP agents offer protection under mixed-film and boundary conditions. (Ref 17).

Both antiwear and EP additives provide protection by a similar mechanism, except that the EP additives typically require higher activation temperatures and loading conditions than the antiwear additives. Simply stated, antiwear additives perform under mild conditions, and EP additives, under severe conditions. The severity of conditions is determined by the "load factor" experienced by the additive and the temperature at which the additive functions. Heavy loading requires EP agents and mild loading requires antiwear agents. Thus, it is important to consider both the load and the temperatures that the equipment is likely to experience before selecting antiwear and EP agents.

Antiwear agents are commonly used in engine oils, automatic transmission fluids, power steering fluids, and tractor hydraulic fluids. Extreme-pressure agents are used in gear oils, other power-transmitting fluids, shock absorber fluids, and metalworking fluids. In gasoline engine oils, the ASTM sequence IIIE and VE engine tests are used to determine the effectiveness of antiwear agents. In gear oils, the CRC L-37 and CRC L-42 axle tests are used to determine the performance of EP agents. The general effectiveness of these additives in metalworking fluids is determined by the use of Timken, four-ball, and Falex tests. Extreme-pressure additives are usually supplemented with antiwear additives to make the formulations effective at lower temperatures and milder loading conditions.

Most antiwear and extreme pressure agents contain sulfur, chlorine, phosphorus, boron, or combinations thereof. The classes of compounds that inhibit adhesive wear include alkyl and aryl disulfides and polysulfides, dithiocarbamates, chlorinated hydrocarbons, and phosphorus compounds (phosphites, phosphates, phosphonates, and dialkyl dithiophosphates).

Both antiwear and EP additives function by thermal decomposition and by forming products that react with the metal surface to form a solid protective layer. This solid metal film fills the surface asperities, thereby reducing friction and

preventing welding and surface wear. The metal film consists of iron halides, sulfides, or phosphates, depending on the antiwear and EP agents used. Friction modifiers differ from antiwear and EP agents in that they form the protective film through physical adsorption instead of chemical reaction.

Many effective extreme pressure and antiwear additives are corrosive to metals. Therefore, lubricants are typically formulated to optimize a balance between extreme-pressure and antiwear protection and corrosivity.

Antiwear Agents. Dithiophosphoric acid zinc salts are the most widely used antiwear agents. In addition to providing antiwear protection, zinc dialkyl dithiophosphates act as oxidation and corrosion inhibitors. They are primarily used in gasoline and diesel engine oils and in industrial lubricants.

Zinc dialkyl dithiophosphates or diaryl dithiophosphates are synthesized by reacting respective dithiophosphoric acids with zinc oxide (Fig. 7). The dithiophosphoric acids are product of the reaction of either an alcohol or a phenol with phosphorus pentasulfide.

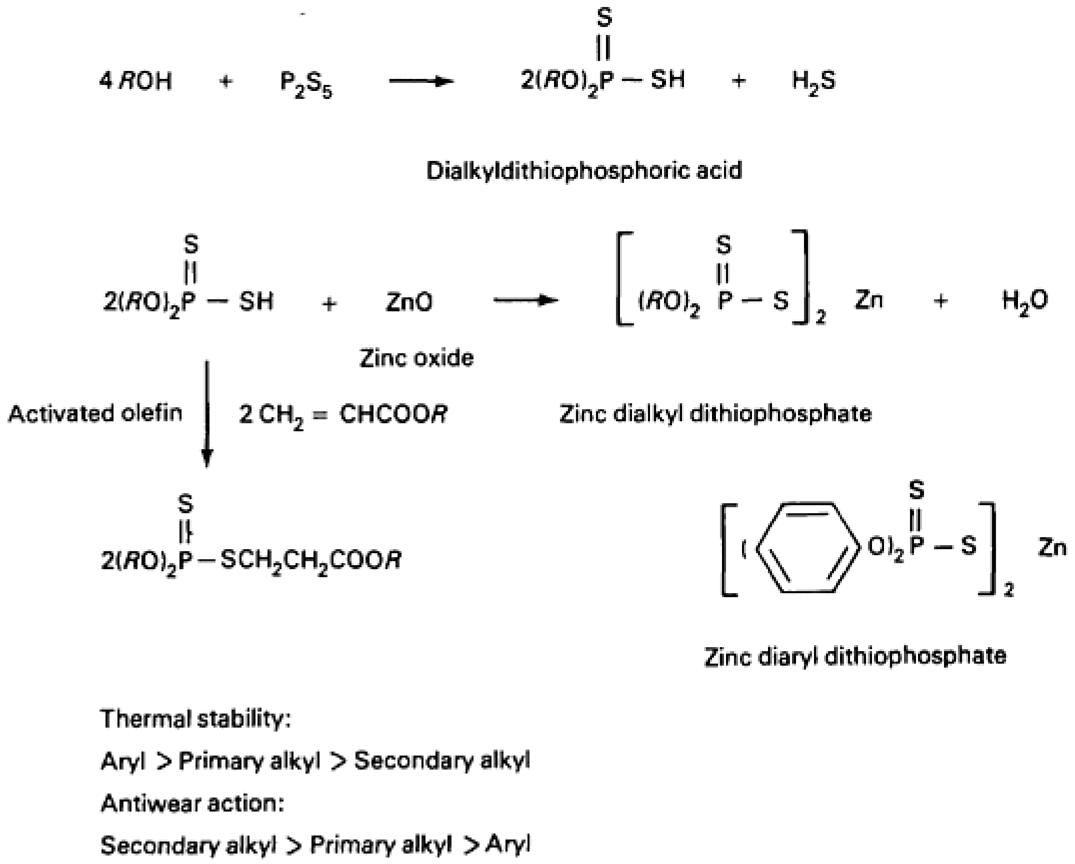


Fig. 7 Preparation of dithiophosphoric acid derivatives

The thermal and hydrolytic stability of these products depends on the nature of the organic group. The dialkyl dithiophosphates derived from primary alcohols are more thermally stable than those derived from secondary alcohols. They are extensively used in formulating gasoline and automotive diesel engine oils. Although the diaryl dithiophosphates are the most thermally stable of this family, they are hydrolytically the least stable. They are not very effective and therefore not often used as antiwear agents. Antiwear performance depends on the thermal stability of the dialkyl dithiophosphate, which in turn depends on whether the alcohol is primary or secondary. Dithiophosphate derivatives decompose (generally below 200 °C, or 390 °F) to form thiols, olefins, polymeric alkyl thiophosphates, and hydrogen sulfide (Ref 18, 19).

Antiseize additives are a separate class of antiwear additives that perform independently of temperature. They improve boundary lubrication by forming a protective film through deposition. Molybdenum disulfide and graphite are common examples that are generally used in greases, some industrial oils, and various break-in lubricants.

Extreme-Pressure Agents. Alkyl and aryl disulfides and polysulfides, dithiocarbamates, chlorinated hydrocarbons, dialkyl hydrogen phosphites, and salts of alkylphosphoric acids are common EP agents.

The polysulfides are synthesized from olefins by reacting with either sulfur or sulfur halides, followed by dehydrohalogenation. Sulfurization of olefins with elemental sulfur or sulfur and hydrogen sulfide yields organic sulfides and polysulfides (Ref 20). Dialkyldithiocarbamate derivatives are prepared either by the neutralization of dithiocarbamic acid (resulting from the low-temperature reaction of a dialkylamine and carbon disulfide) with bases or by the addition of this acid to activated olefins such as alkyl acrylates (Ref 21). The synthesis of these materials is presented in Fig. 8.

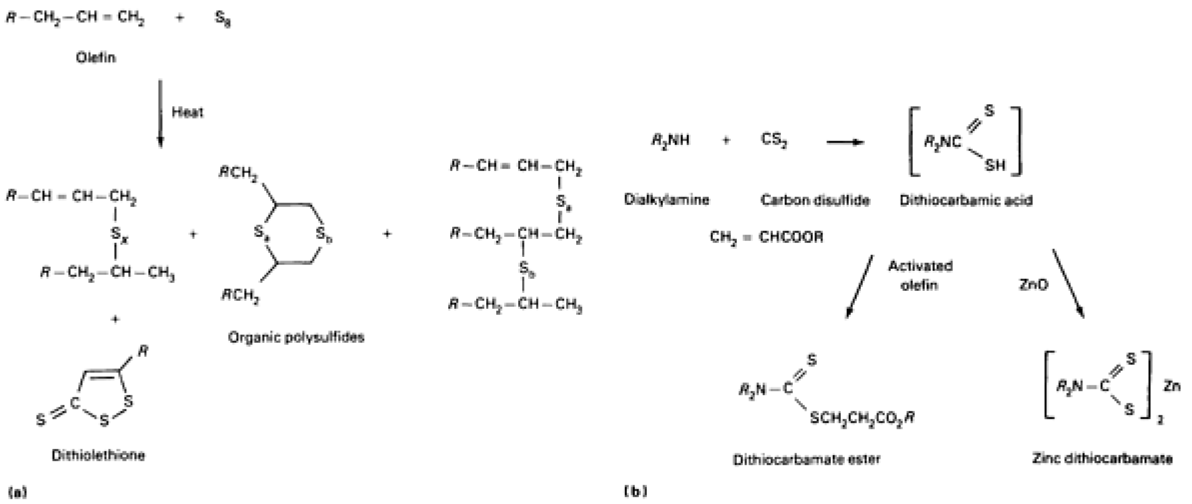


Fig. 8 Synthesis of sulfur EP agents. (a) Olefin sulfurization. (b) Dithiocarbamic acid derivatives

The phosphites are obtained either by reacting an alcohol or phenol with phosphorus trichloride or by a transesterification reaction (Ref 22). Alcohols and phenols react with phosphorus pentoxide to yield a mixture of alkyl(aryl)phosphoric acid and dialkyl(diaryl)phosphoric acid (Ref 23). These acids form salts, upon treatment with bases. Phosphates can also be prepared by the oxidation of phosphites. The preparation of these materials is outlined in Fig. 9.

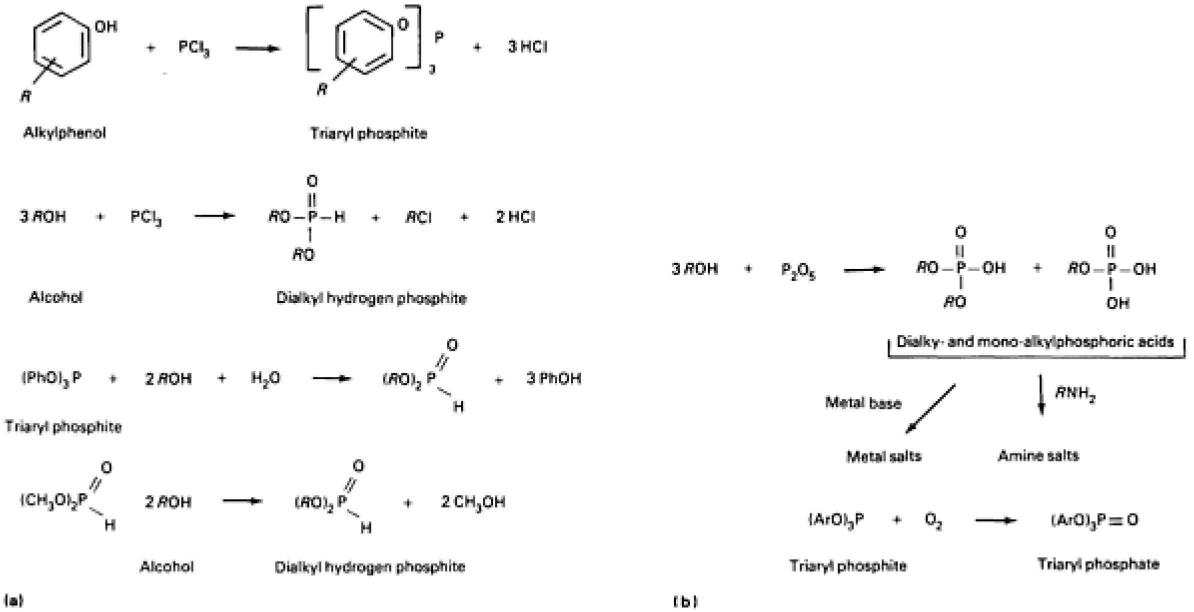


Fig. 9 Synthesis of phosphorus EP agents. (a) Phosphites. (b) Phosphates

Disulfide and polysulfides decompose to form a protective sulfide layer on the metal surfaces at temperatures above 200 °C (390 °F). The thickness of this layer depends on the quantity and lability of the sulfur in the additive. Sulfurized fatty oils and sulfurized olefins are the most commonly used products in this class.

Chlorine compounds provide protection through the formation of a metal chloride film under boundary lubrication conditions. A detrimental aspect of chlorine-based EP agents is the formation of hydrogen chloride in the presence of moisture, which can cause severe corrosion problems. Chlorinated paraffins with 40 to 70% chlorine by weight were popular. Environmental concerns about the negative effects of chlorine are minimizing the use of these additives.

Phosphorus compounds react with the metal surface to make a metal phosphite or phosphate protective film. This film is formed at a much higher temperature than that formed by the sulfur EP agents. Tricresyl phosphate is the best-known example of the phosphorus EP agents. Dialkyl hydrogen phosphites and phosphonic and phosphoric acid salts are further examples of these agents.

Friction Modifiers/Antisquawk Agents

Lubrication is necessary to facilitate the countermovement of two sliding surfaces. This function, which is usually performed by mineral oil, can be made more efficient by using friction modifiers. These agents control friction and thus prevent scoring, and reduce wear and noise, as well.

Friction modifiers are long-chain molecules with a polar end group and a nonpolar hydrocarbon chain. The polar end group physically adsorbs itself on the metal surface, whereas the hydrocarbon chain increases the strength of lubricant film through association. Under heavy load conditions, the EP agent replaces the friction modifier and takes over the function of damage prevention. The process is repeated again as the load eases off. Friction modifiers have a finite life related to their oxidative and thermal stability.

Friction modifiers are commonly used in gasoline engine oils, automatic transmission fluids, tractor hydraulic fluids, power steering fluids, shock absorber fluids, and metalworking fluids. In passenger car applications with federal government mandated fuel economy, the lubricant suppliers use these additives as a competitive marketing tool. In automatic transmission fluids and limited slip axle lubricants, friction modifiers are used to control torque application through clutch and band engagements. Antisquawk additives, which are functionally similar to friction modifiers, are used to reduce objectionable mechanical noises, such as squawk and chatter. Antisquawks are primarily used in automatic transmission and tractor hydraulic fluids, automotive gear oil, and industrial oils. This class of additives includes fatty

alcohols, fatty acids, fatty amides, and molybdenum compounds and graphite. Friction modifier and antisquawk performance specifications include evaluation of the product's load-carrying capacity.

In the case of fatty alcohol and fatty acid families, the friction-modifying properties are a function of the length and nature of the hydrocarbon chain and the nature of the functional group. Long and linear-chain materials are better friction modifiers than short and branched-chain materials. Fatty acids are better than fatty amides, which in turn are better than fatty alcohols. Saturated acids that contain a 13 to 18 carbon chain are generally preferred. Lower molecular weight fatty acids are avoided because of their corrosivity.

Fatty acid derivatives are the most commonly used friction modifiers. Because graphite presents problems when used in conventional lubricants, its use is currently limited to greases.

Oxidation Inhibitors

All modern lubricants, by virtue of being hydrocarbon based, are susceptible to oxidation (Ref 24, 25). Each type of base stock (mineral or synthetic) has a stable threshold, beyond which stabilizers or oxidation inhibitors are needed to retard oxidation.

Adequately refined base oils contain some sulfur- and nitrogen-base "natural" inhibitors. Under mild conditions, these inhibitors are sufficient to protect lubricants against oxidation. However, most modern uses require supplemental inhibitors to protect lubricants under increasingly hostile conditions.

Most lubrication applications expose lubricants to oxygen in some manner. The attack of oxygen on the weaker sites in the lubricant, predominantly allylic, benzylic, and tertiary hydrogens, produces peroxy or other radicals.

The process of oxidation proceeds in three stages. During the initiation stage, the lubricant and the fuel react with oxygen to form radicals. During the propagation stage, these radicals react with oxygen and the lubricant to form hydroperoxides. The hydroperoxides decompose to form a variety of additional radicals and oxygen-containing compounds such as alcohols, ketones, and carboxylic acids. The carboxylic acids attack iron metal and copper and lead bearings to form metal carboxylates, which further enhance the rate of oxidation. During the termination stage, the radicals either self-terminate or terminate by reacting with oxidation inhibitors (Ref 24). The mechanism of oxidation is illustrated in Fig. 10, whereas Fig. 11 depicts the mechanism of the hydroperoxide decomposition.

Initiation:

$RH + O_2$	$\longrightarrow$	$R^{\cdot} + HO_2$	(1)
$RH + O_2 + RH$	$\longrightarrow$	$R^{\cdot} + H_2O + R^{\cdot}$	(2)

Propagation:

$R^{\cdot} + O_2$	$\longrightarrow$	$ROO^{\cdot}$	(3)
$ROO^{\cdot} + RH$	$\longrightarrow$	$ROOH + R^{\cdot}$	(4)

Peroxide Decomposition

$ROOH$	$\longrightarrow$	$RO^{\cdot} + {}^{\cdot}OH$	(5)
$2ROOH$	$\longrightarrow$	$RO^{\cdot} + ROO^{\cdot} + H_2O$	(6)
$RO^{\cdot} + ROOH$	$\longrightarrow$	Various Products	(7)
$\left[\begin{array}{l}ROO^{\cdot}\\ ROOH\end{array}\right.$	$\longrightarrow$	Non-radical products	(8)
$X + ROOH$	$\longrightarrow$	Free radicals	(9)
$Y + ROOH$	$\longrightarrow$	$ROH + YO$	(10)
$Z + ROOH$	$\longrightarrow$	Inactive products + Z	(11)
$M + ROOH$	$\longrightarrow$	Free radicals	(12)

Termination:

$ROO^{\cdot} + ROO^{\cdot}$	$\longrightarrow$	Inactive products	(13)
$ROO^{\cdot} + IH$	$\longrightarrow$	$ROOH + I^{\cdot}$	(14)

RH = Organic substrate
$ROO^{\cdot}$ = Peroxy radical
ROOH = Hydroperoxide
X,Y,Z = Decomposers of organic peroxides
M = Metal
$I^{\cdot}$ = Stable or unreactive radical
IH = Chain breaking inhibitor

Fig. 10 Mechanism of oxidation. Source: Ref 24

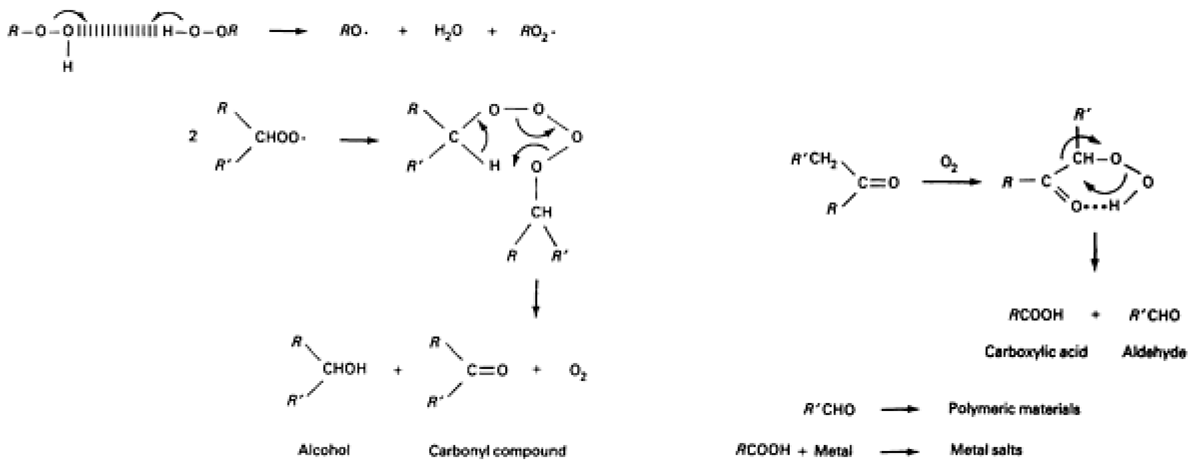


Fig. 11 Hydroperoxide decomposition

An increase in temperature affects the oxidation process most profoundly, with the rates of oxidation approximately doubling with every 10-degree rise in temperature. Wear metals can further enhance the rate of oxidation (Ref 26). If this process is not controlled, the lubricant decomposition will lead to oil thickening, sludge formation, and the formation of varnish, resin, and corrosive acids (Ref 3, 4). Oil thickening occurs primarily because of polymerization or association of certain oxidation products.

Oxidation inhibitors function by circumventing the radical chain mechanism of the oxidation process. Oxidation inhibitors can be classified as hydroperoxide decomposers and radical scavengers, depending on the mode of their controlling action. Sulfur- and phosphorus-containing inhibitors, such as sulfides, dithiocarbamates, phosphites, and dithiophosphates, act as hydroperoxide decomposers. Nitrogen- and oxygen-containing inhibitors, such as arylamines and phenols, act as radical scavengers (Ref 24, 25).

The hydroperoxide decomposers convert the chain-propagating hydroperoxides to alcohols while getting themselves oxidized to higher oxidation levels. The radical scavengers remove the radicals from the oxidation process by transferring a hydrogen atom to the radical, thereby making it innocuous. In the process, these materials do get converted into radicals, but unlike the radicals from the oxidation process, these radicals are incapable of propagating oxidation. The reason is that these new radicals are either too sterically hindered or delocalized to take part in the oxidation process. The mechanisms by which the hydroperoxide decomposers and radical scavengers perform are summarized in Fig. 12 and 13, respectively.

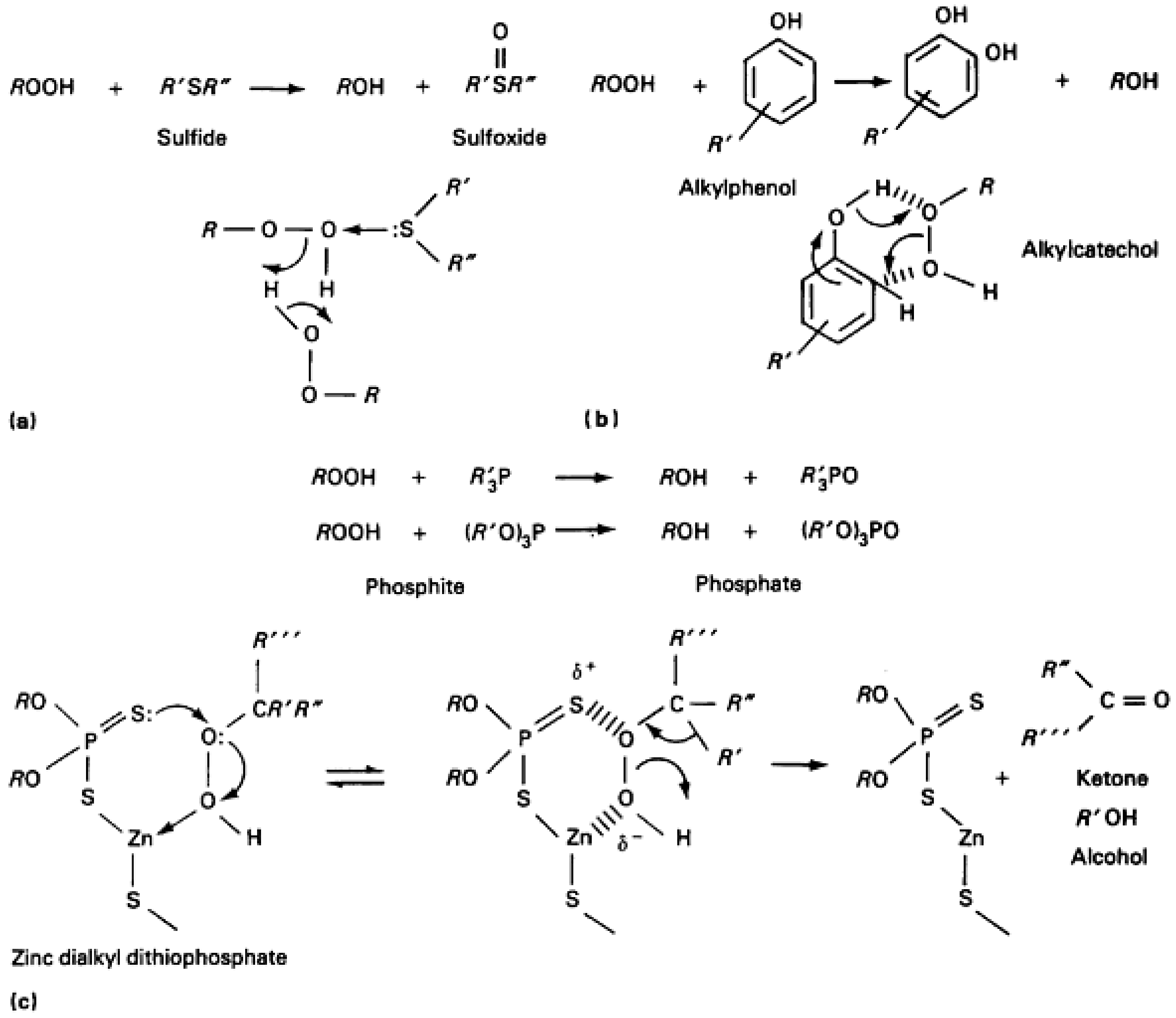
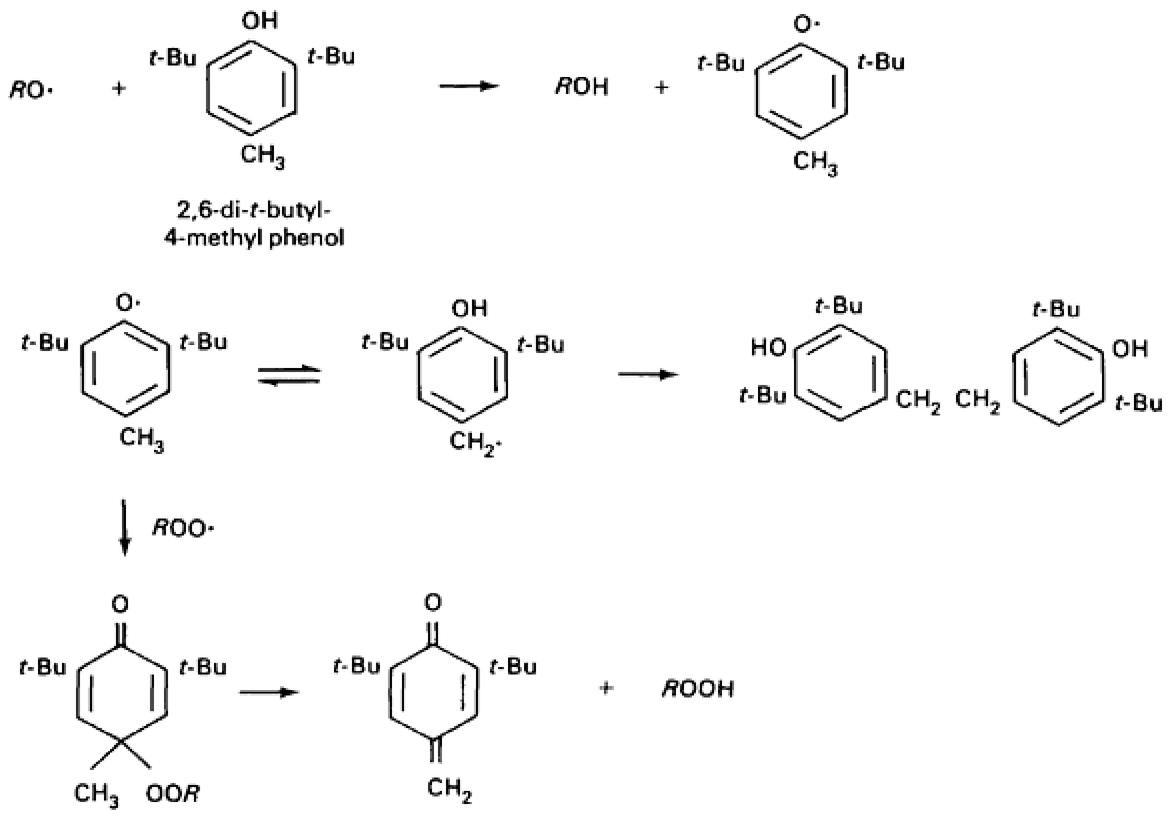
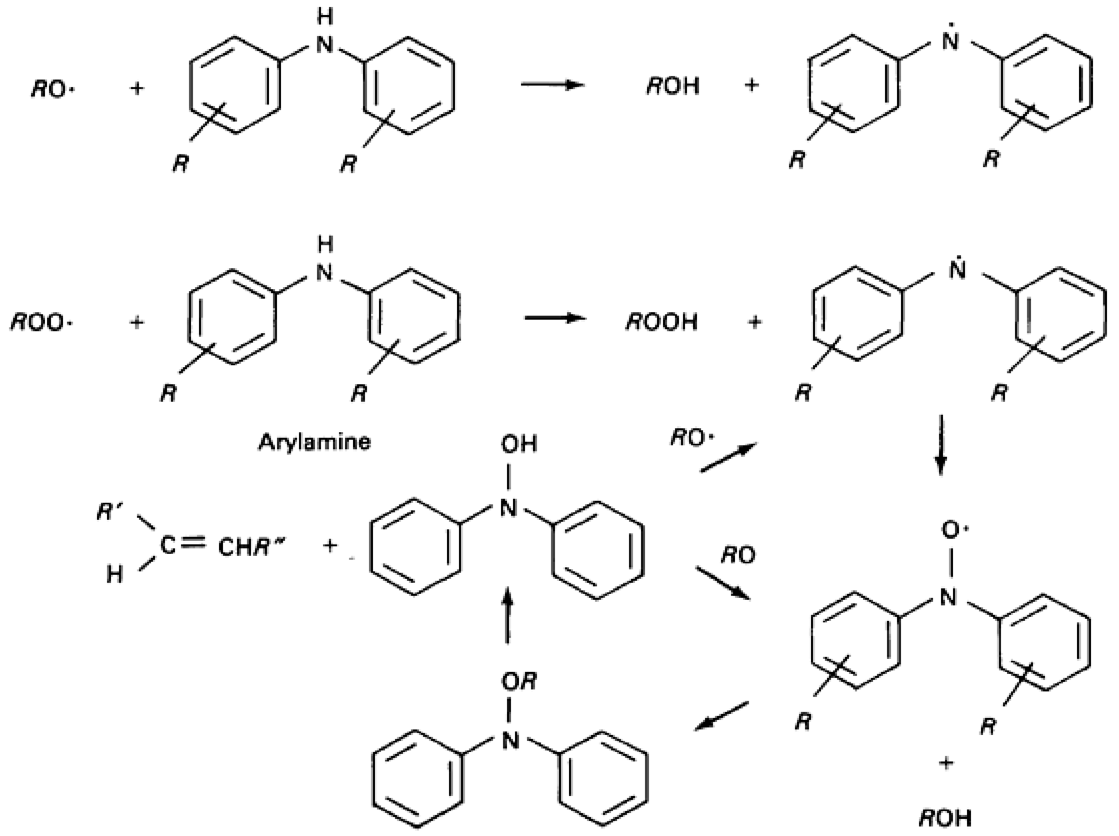


Fig. 12 Hydroperoxide decomposers and mode of action. (a) Sulfur compounds. (b) Phenols. (c) Phosphorus compounds



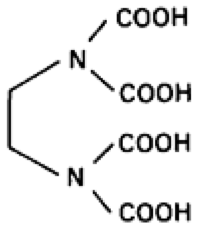
(a)



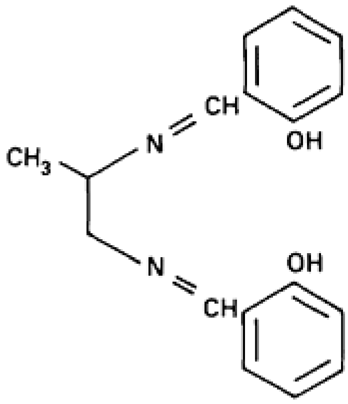
(b)

Fig. 13 Radical scavengers and mode of action. (a) Hindered phenols. (b) Arylamines

Ions of heavy metals, such as iron or copper, are known to promote oxidation (Ref 27). Metal deactivators, another class of oxidation inhibitors, slow down the oxidation process by forming complexes with these metal ions and taking them out of the chain reaction. Ethylenediam-inetetraacetic acid derivatives and N,N-disalicylidene-1,2-propanediamine fall under this class (Fig. 14). Sometimes, compounds that intercept oxidation by different mechanisms reflect synergism when present together. Synergism is an effect greater than the additive effect of the two or more compounds (Ref 28).



Ethylenediaminetetraacetic acid



N,N-disalicylidene-1,2-propanediamine

Fig. 14 Metal deactivators

Oxidation inhibitors are used in almost all lubricants. Gasoline and diesel engine oils and automatic transmission fluids account for 60% of the total use. High operating temperature and high air exposure applications require a high level of oxidation protection. Zinc dialkyl dithiophosphates are the primary inhibitor type, followed by aromatic amines, sulfurized olefins, and phenols.

A number of tests are used to assess the oxidation stability of a lubricant under conditions of accelerated oxidation. The ASTM sequence IIIE (viscosity increase), sequence VE (sludge and varnish formation), and CRC L-38 (bearing corrosion) tests are used for engine oils. The CRC L-60 test is used for gear oils, and ASTM D 943 and ASTM D 2272 tests are used for turbine oils.

Rust and Corrosion Inhibitors

Rust and corrosion represent the damage done to metal surfaces by the attack of atmospheric oxygen and acidic products. The rust and corrosion rate, which is usually low, increases greatly in the presence of water and polar impurities. Internal combustion engines contain all the elements necessary to promote these types of surface damage. Air is entrained in the oil and fuel, and water and organic acids form during the combustion and decomposition processes. Rust and corrosion inhibitors provide a barrier between the metal surface and these harmful elements. These inhibitors are of two types: those that neutralize acids and those that form protective films. The film formers attach themselves strongly to the metal surface and form an impenetrable protective film. The film formation can occur through either physical adsorption or chemical reaction. The first mechanism involves the formation of densely packed protective layers of a somewhat transient nature and the second, a more permanent protective film resulting from a chemical reaction between the inhibitor and the metal surface.

Long-chain organic molecules, such as alkanolamines and saturated and unsaturated fatty acid amides, are examples of physical adsorbers. Phosphoric acid, dithiophosphoric acid, and succinic acid derivatives are examples of chemical film formers. Basic detergents are excellent rust and corrosion inhibitors, because they provide protection both by neutralizing and by forming physically adsorbed films. For many applications, rust and corrosion inhibitor systems are required to provide both the vapor phase and the liquid phase protection, that is, for surfaces above and below the lubricant level.

Long-chain amines, basic sulfonates, carboxylic acid derivatives, and thiadiazole and triazole derivatives are the most commonly used rust and corrosion inhibitor types. These additives have major uses in engine oils, gear oils, metalworking fluids, and greases. Thiadiazole and triazole derivatives are especially useful in protecting against nonferrous or yellow metal corrosion. Figure 15 depicts common types of rust and corrosion inhibitors.

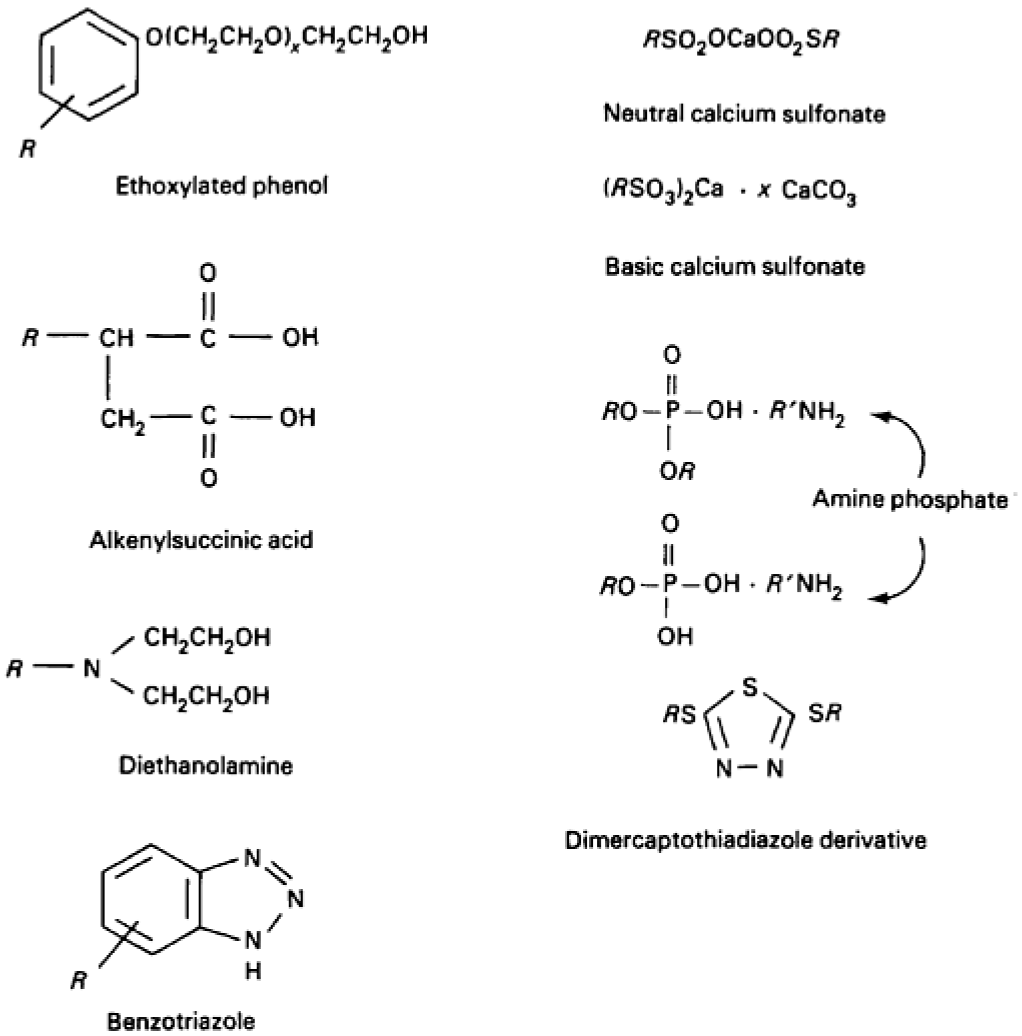


Fig. 15 Corrosion inhibitors

A number of ASTM-specified tests are used for hydraulic and metalworking fluids. For engine oils, the ASTM sequence IID engine test is used to assess the corrosion-inhibiting ability of the lubricant.

Emulsifiers and Demulsifiers

Emulsifiers are chemical compounds that enable two immiscible fluids to form an intimate mixture known as an emulsion. Water and oil mixtures are often used as lubricants in a number of industries for a variety of applications. Such lubricants are low cost, easy to dispose of, and have fire-retardant properties. Emulsions of water and mineral oil are primarily used in metal-working and hydraulic applications. Emulsions have to possess a number of desirable properties.

They should be stable over long periods of time, must possess good lubricating properties, should not attack seals and metals, and should be easy to demulsify for disposal.

Certain lubricant formulations, in the presence of water, have an increased tendency to form emulsions. This is due to the presence of chemical additives (such as detergents) that act as surfactants. Demulsifiers are added to such formulations to enhance water separation and suppress foam formation.

Emulsifiers and demulsifiers are basically surfactants and contain hydrophilic and hydrophobic moieties. The hydrophilic moiety is a hetero atom (N,O,S and P)-based polar functional group, which is attached to a hydrophobic portion, a hydrocarbon chain of sufficient length to provide proper solubility or dispersibility in the oil phase. The additives that belong to this group can be classified as being non-ionic or ionic, depending on whether the polar part is uncharged or charged. The ionic compounds can be subcategorized as cationic if the charge is positive, and anionic if the charge is negative (Ref 29). The generalized structures for emulsifiers and demulsifiers are given in Fig. 16.

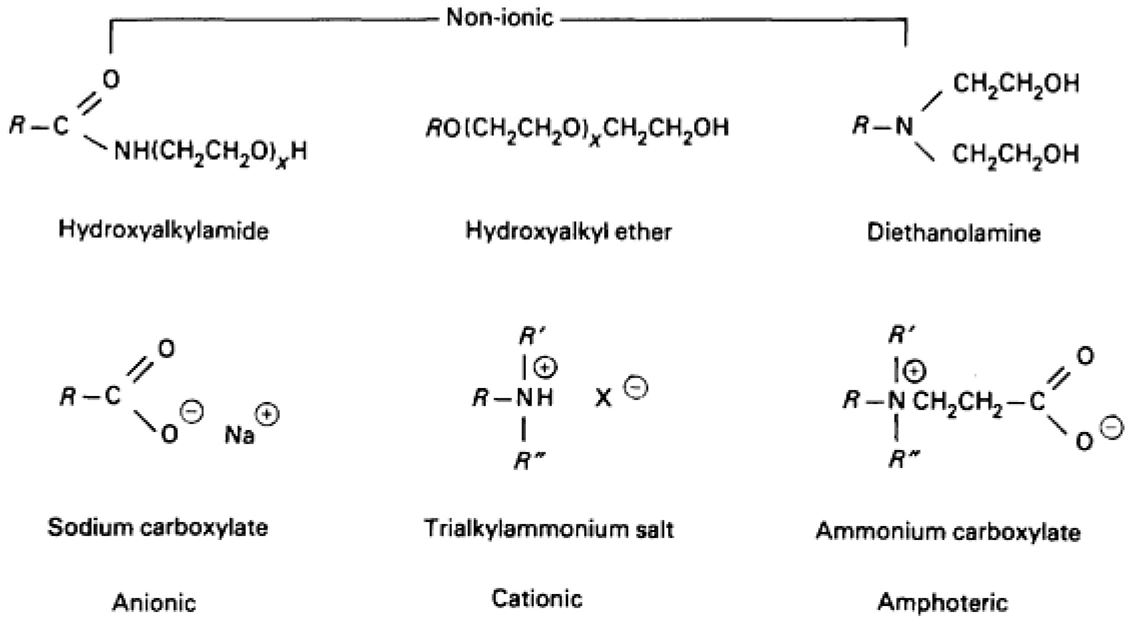


Fig. 16 Emulsifiers and demulsifiers

It is important to note that only the charge on the functional group attached to the carbon chain is used in this classification. The charge on the counterion, which is usually inorganic in origin, is ignored. The term amphoteric applies to a group of additives that contains both the cationic and anionic groups of organic origin. They possess the structural features and properties of both the cationic and anionic materials grouped together.

Emulsifiers reduce the surface tension of water, and thereby facilitate thorough mixing of oil and water to form an emulsion. The efficiency of the emulsifier depends on its molecular weight (usually less than 2000), its hydrophile-lipophile balance (HLB) value, both the pH and hardness of water, the nature of the oil, and operating conditions, such as temperature. The emulsifiers with an HLB value from 3 to 6 are suitable for water-in-oil emulsions, and those with an HLB value from 8 to 18 are suitable for oil-in-water emulsions.

Demulsifiers perform the opposite function and enhance water separation. Structurally, most demulsifiers are oligomers or polymers with a molecular weight of up to 100,000 and contain from 5 to 50% polyethylene oxide in a combined form. Demulsifiers are commonly block copolymers of propylene oxide or ethylene oxide, combined with initiators such as glycerol, phenol-formaldehyde resins, siloxanes, polyamines, and polyols (Ref 30). Lower molecular weight materials, such as alkali metal or alkaline earth metal salts of dialkylnaphthalenesulfonic acid, are also useful in lubricant-related applications. Polymers containing from 20 to 50% ethylene oxide are suitable for water-in-oil emulsions. These materials concentrate at the water-oil-interface and create low-viscosity zones, thereby promoting droplet coalescence and gravity-driven phase separation.

As a general rule, non-ionic emulsifiers are used in metalworking fluids based on naphthenic stocks, and fatty acid carboxylates in those based on paraffinic stocks. Polyalkylene glycols (hydroxyalkyl ethers) are sometimes avoided because their enhanced solubility in water does not allow clean separation for disposal.

Polyethylene oxide derivatives and salts of carboxylic and sulfonic acids are the most commonly used emulsifiers. Their major use is in metalworking. Demulsifiers are used in applications where water contamination of the lubricant is a problem and quick separation of water is desired. Automatic transmission fluids, hydraulic fluids, and industrial gear oils are examples of such lubricants.

Pour-Point Depressants

The pour point is the lowest temperature at which a fuel or an oil will pour when cooled under defined conditions. In general, the pour point is indicative of the amount of wax (straight-chain paraffins) in an oil. At low temperatures wax tends to separate as crystals with a lattice-type structure. These crystals can trap a substantial amount of oil via association, inhibit oil flow, and ultimately hinder proper lubrication of the equipment.

Base oil suppliers remove most of the wax during petroleum refining. However, complete dewaxing of the base oils is not practical because of process limitations, economics, and the desirable presence of wax, which offers a high viscosity index character (Ref 16). Additives called pour-point depressants are used to enable mineral oils to function efficiently at low temperatures. Current practice favors mild dewaxing in combination with the use of pour-point depressants.

A good pour-point depressant can lower the pour point by as much as 40 °C (70 °F). These additives are commonly used in applications that require the use of mineral oil, usually below 0 °C (32 °F). Pour-point depressants have virtually no effect on either the temperature at which the wax crystals start to precipitate (cloud point) or the amount of wax that separates. They essentially act as wax-crystal modifiers and function by altering the crystal size. They do this either by absorption onto the surface of newly formed crystals or by co-crystallizing with the precipitating wax. Both mechanisms inhibit lateral crystal growth and keep the bulk oil fluid. Of the commercial pour-point depressants, alkylaromatics are believed to perform via the absorption mechanism and aliphatic polymers via co-crystallization (Ref 31).

The molecular weight and structure of polymeric pour-point depressants enable them to be effective over a wider range than their low molecular weight counterparts. The extended range of performance, in the case of polymers, has been explained on the basis of their limited solubility in petroleum fractions. As the temperature decreases, different polymer segments become successively co-crystallizable.

A good pour-point depressant should possess one or more of the following structural features:

- Polymeric structure
- Waxy and nonwaxy components
- Comb structure (backbone with pendant groups
- Broad molecular weight distribution

Most commercial pour-point depressants are organic polymers, although some nonpolymeric substances have been shown to be effective. Tetra(long chain)alkyl silicates, phenyltristearyloxysilane, and pentaerythritol tetrastearate are examples of the nonpolymeric type. Commercial pour-point depressants include alkylated bicyclic aromatics, styrene esters, polymethacrylates, polyfumarates, oligomerized alkylphenols, dialkaryl esters of phthalic acid, ethylene-vinyl acetate copolymers, and other mixed hydrocarbon polymers. Figure 17 contains structures of common pour-point depressants. Polymethacrylates and polyfumarates are the most commonly used chemical types. High molecular weight polymethacrylate derivatives can act both as viscosity improvers and pour-point depressants. When this chemistry is used for viscosity improvement, the need for a pour-point depressant is minimized.

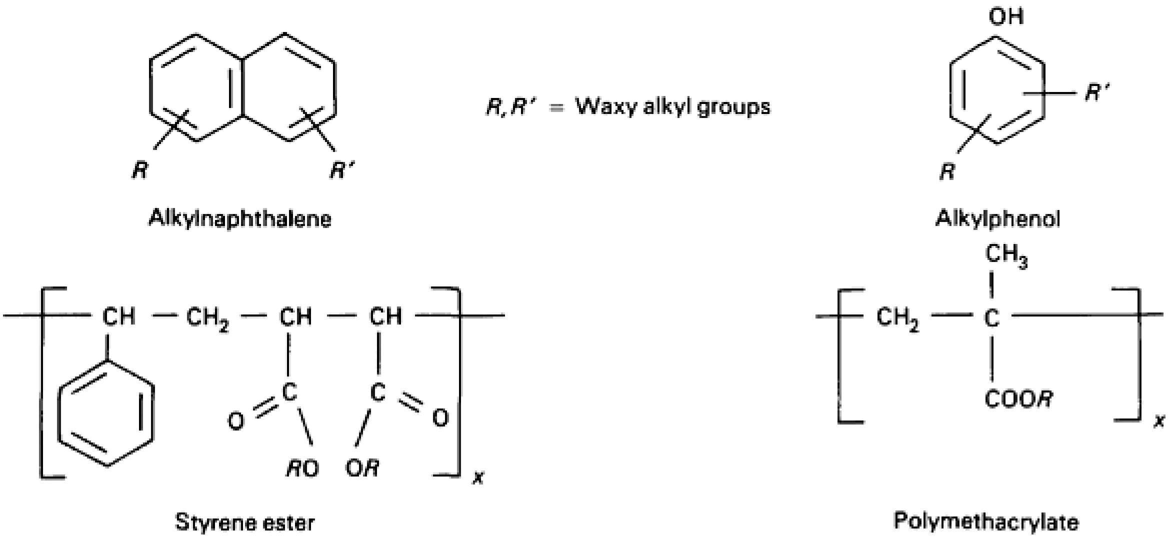


Fig. 17 Pour-point depressants

Pour-point depressants are used at treatment levels of 1% or lower. In nearly all cases, there is an optimum concentration above and below which the pour-point depressants become less effective.

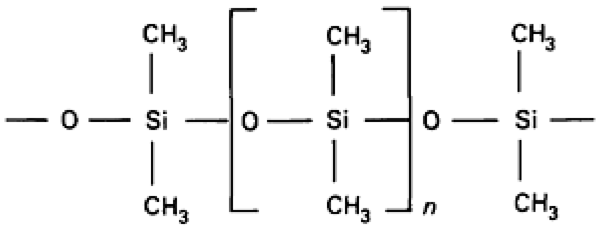
Pour-point depressants are used in crankcase engine oils, automatic and power transmission fluids, automotive gear oils, tractor fluids, hydraulic fluids, and circulating oils. The performance of pour-point depressant is determined in each base stock by using one or more of the following tests: ASTM D 97, pour point of petroleum oil; ASTM 3829, borderline pumping; and ASTM D 2602, apparent viscosity at low temperature using cold-cranking simulation.

Foam Inhibitors

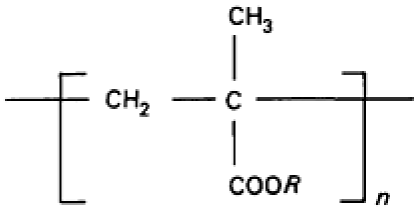
Almost every lubricant application involves some kind of agitation, which encourages foam formation through air entrainment. Excessive foaming will result in ineffective lubrication and, over time, will cause oxidative degradation of the lubricant. The viscosity of a lubricant and surface tension determine the stability of the foam. Low-viscosity oils produce foams with large bubbles that tend to break quickly. High-viscosity oils, on the other hand, generate stable foam that contain fine bubbles and are difficult to break. The presence of surface-active materials, such as dispersants and detergents, further increases the foaming tendency of the lubricant.

Foam inhibitors inhibit foam formation by altering the surface tension of the oil and by facilitating the separation of air bubbles from the oil phase. These additives generally have a limited solubility in oil, and therefore are added as very fine dispersions. Foam inhibitors are effective at very low levels (3 to 150 parts per million).

Silicones (polysiloxanes or dimethylsiloxanes) and polyacrylates are commonly used as foam inhibitors, with silicones being more popular. ASTM D 892 is used to assess the foaming tendency of a lubricant. Figure 18 shows the structures for the two types of additives.



Polydimethylsiloxane



Alkyl polymethacrylate

Fig. 18 Foam inhibitors

Viscosity Improvers

Mineral oils, which are effective lubricants at low temperatures, become less effective lubricants at high temperatures. At high temperatures, their film-forming ability (in the hydrodynamic lubrication regime) diminishes, because of a drop in viscosity. Prior to the use of viscosity improvers and the introduction of multigrade oils, this problem was partly overcome through seasonal oil changes.

The principal function of a viscosity improver is to minimize viscosity variations with temperature. Previously, viscosity index was used as a measure of the response of an oil to temperature changes. Viscosity index, which is derived from the viscosity of the oil at 40 °C (105 °F) and at 100 °C (212 °F), is no longer as meaningful. This is because modern equipment is operating at extreme temperatures (-40 to 150 °C, or -40 to 300 °F) and, at these temperatures, the viscosities do not conform to those predicted by the viscosity index (Ref 31).

Viscosity improvers are typically added to a low-viscosity oil to improve its high-temperature lubricating characteristics. These are organic polymers that minimize viscosity change with a change in temperature. This represents a practical means by which the operating range of mineral oils is extended to high temperatures without adversely affecting their low-temperature fluidity.

At low temperatures, the polymer molecules occupy a small volume (hydrodynamic volume) and therefore have a minimum association with the bulk oil. The effect is little or no viscosity increase. However, the situation is reversed at high temperatures because polymer chains extend or expand as a consequence of added thermal energy. This increases the association of the polymer with the bulk oil (Ref 16), because of an increase in surface area. The result is an effective increase in viscosity. Figure 19 illustrates oil thickening by viscosity improvers.

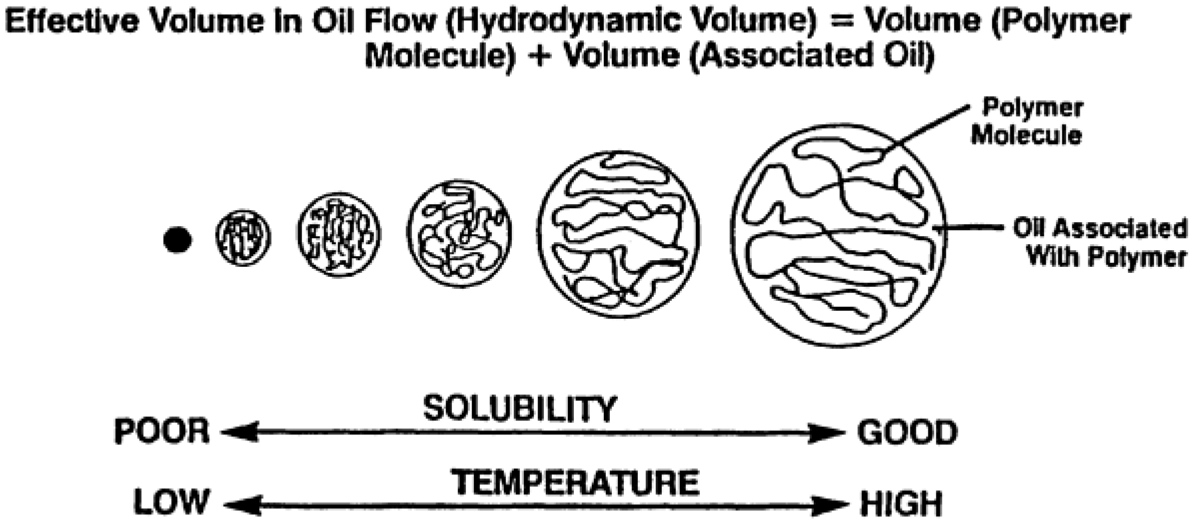


Fig. 19 Viscosity improvers and mechanism of thickening. Source: Ref 16

It is important to note that the superior viscosity-temperature relationship (as expressed by the viscosity index) of multigrade oils is not solely due to greater preferential swelling of the polymer at higher temperatures than at lower temperatures. The good viscosity-temperature behavior of low-viscosity oils that are used to prepare multigrade oils makes a substantial contribution toward the overall effect (Ref 32).

Olefin copolymers (OCP), polymethacrylates (PMA), hydrogenated styrene-diene (STD), and styrene-polyester (STPE) polymers are the common types of viscosity improvers used in modern lubricants. Olefin copolymers are rubber-like materials prepared from ethylene and propylene mixtures through vanadium-based Ziegler-Natta catalysis. The ethylene-to-propylene ratio and their proper distribution in the backbone are critical to the low-temperature properties of the polymer. The OCP products are inferior to polymethacrylates in low-temperature properties and generally require the use of pour-point depressants to improve the low-temperature performance of lubricants containing them.

Styrene-diene polymers are produced by an-ionic polymerization of styrene and butadiene or isoprene. This type of polymerization produces polymers with a narrower molecular weight distribution than those obtained by the use of Ziegler and Friedel-Crafts catalysts, or free radical initiators. These polymers, therefore, possess the best thickening power of the types discussed so far. However, the presence of aromatic rings and double bonds makes these polymers more susceptible to oxidation. This problem is somewhat overcome by catalytically hydrogenating the double bonds.

A different type of polymer, called a star polymer, with a thickening power similar to that of the styrene-diene type, but a better shear stability, has recently become commercially available. A clustered polyanion (prepared from divinylbenzene and styrene) forms the center of the star, and the rays are composed of polymerized diene monomers. Figure 20 depicts the synthesis of olefin and styrene-diene copolymers.

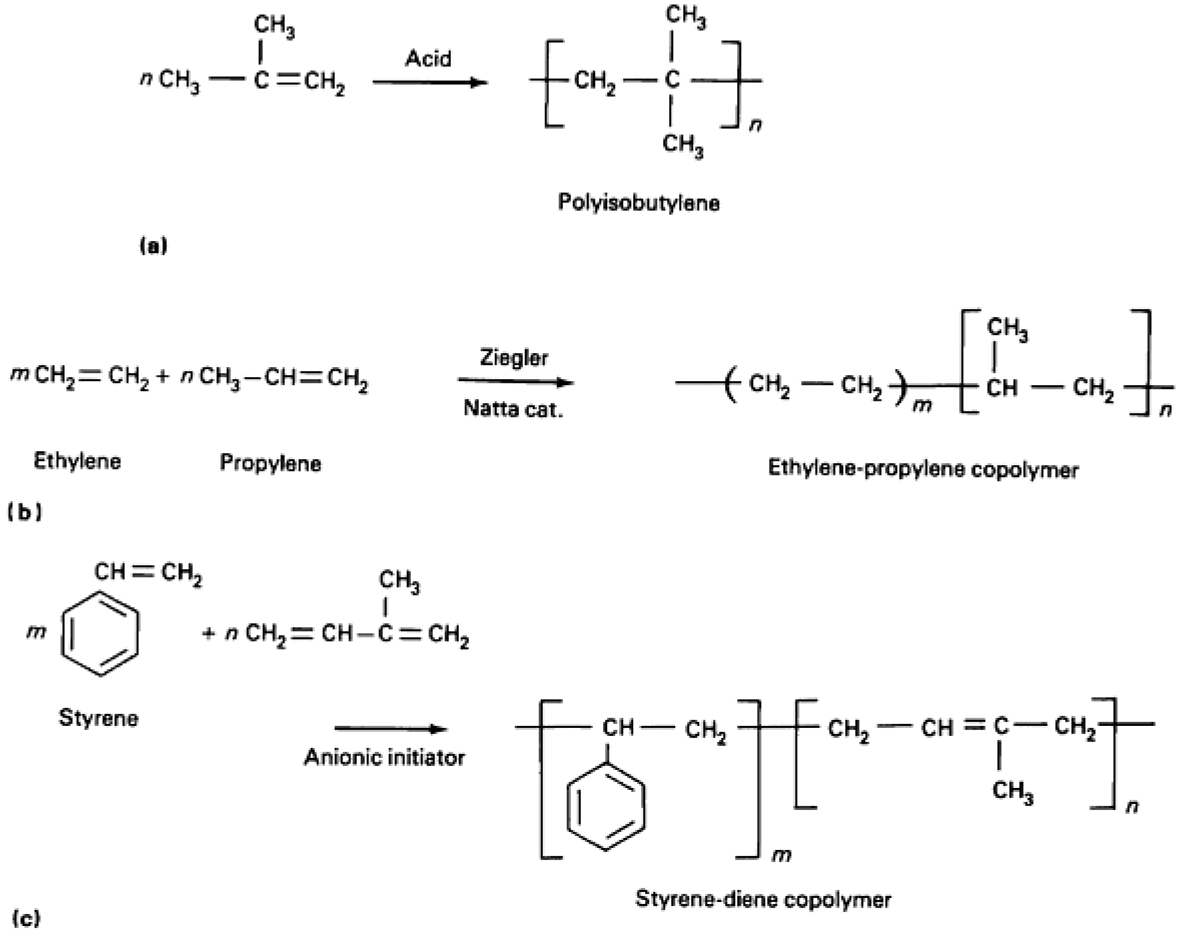


Fig. 20 Synthesis of polyolefins (a), olefin copolymers (b), and styrene-diene copolymers (c)

Polymethacrylates are produced by free radical polymerization of alkyl methacrylates. The alkyl group can be altered so as to obtain a product with the best oil solubility and the optimum viscosity-improving properties. Free radical polymerization produces polymers with a relatively broad molecular weight distribution, thereby reducing their thickening efficiency. Although polymethacrylates are superior to olefin copolymers in being oxidatively and thermally more stable and possessing better low-temperature properties, they have only moderate viscosity-improving ability on an equal molecular weight basis.

Styrene-polyester polymers are prepared by first co-polymerizing styrene and maleic anhydride and then esterifying the intermediate using a mixture of alcohols. Normally, the esterification step is taken to about 90 percent or more, followed by postneutralization using a bifunctional or polyfunctional amine. Because of the presence of basic nitrogen, these polymers function both as a dispersant and as a viscosity improver. Hence, they are called dispersant-viscosity improvers. Figure 21 summarizes the methods of synthesis for the polyacrylate and styrene-polyester type viscosity improvers.

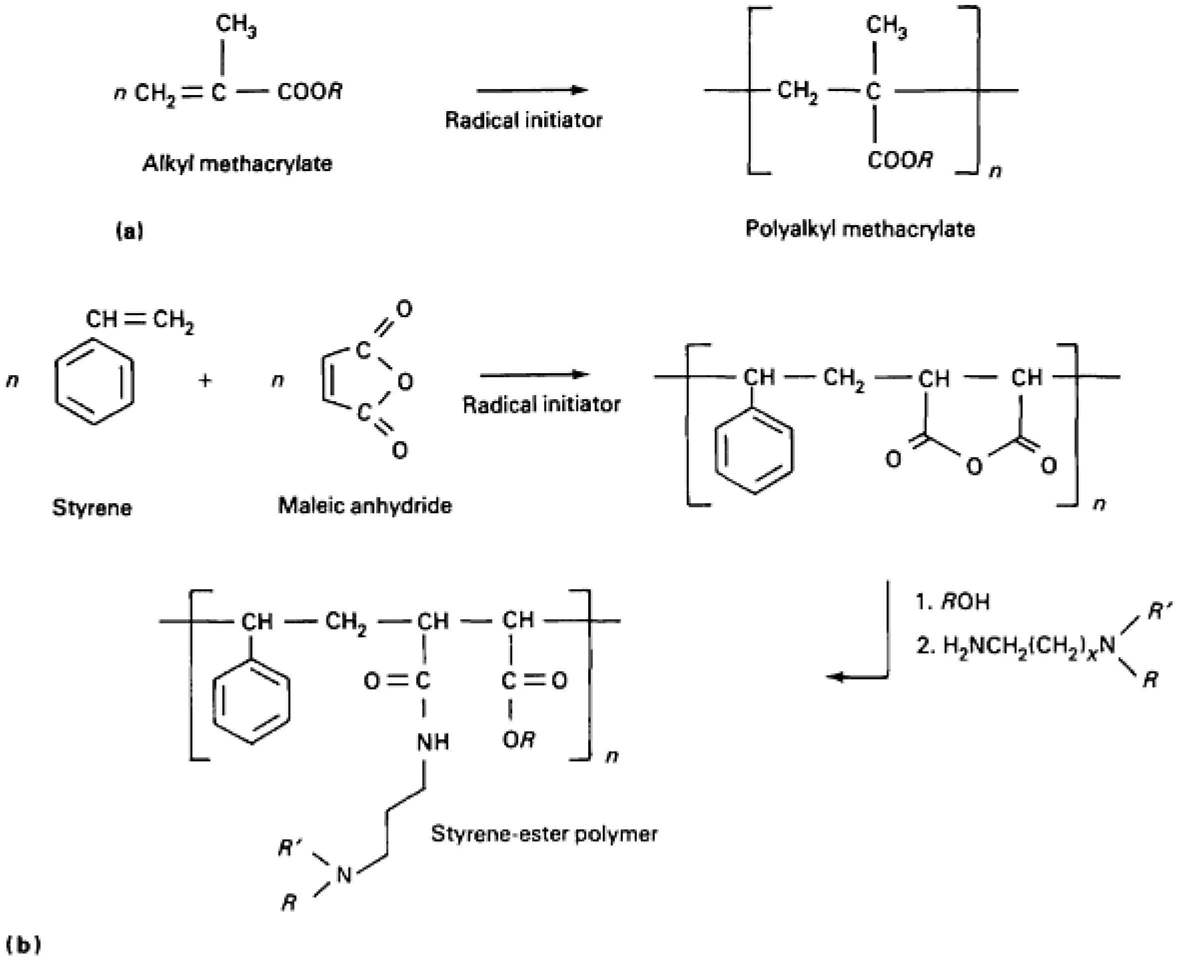


Fig. 21 Synthesis of polyacrylates (a) and styrene-esters (b)

Dispersant-viscosity improvers based on polyacrylate and OCP are also commercially available. Dispersancy is obtained by including basic nitrogen-containing monomers during the polymerization process. N-Vinylpyrrolidinone, vinylpyridines, and dimethylaminoethyl methacrylate are examples of such monomers (Fig. 22).

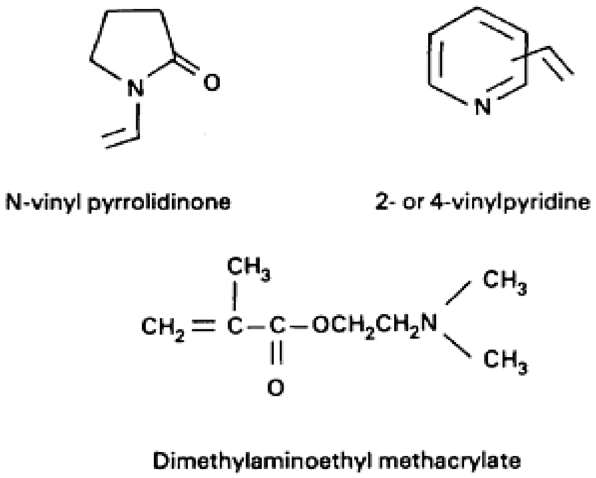


Fig. 22 Nitrogen-containing monomers

Thickening efficiency and shear stability are two important considerations when selecting a polymer for use as a viscosity improver. The thickening efficiency of a polymer is a direct function of its molecular weight. On an equal weight basis, a high molecular weight polymer provides higher viscosity than a low molecular weight polymer. Shear stability, the ability of the polymer to withstand the mechanical shearing forces encountered during use, is inversely related to its molecular weight; the lower the molecular weight, the higher the shear stability. A number of tests are available for measuring the viscosity-improving properties and the shear stability of polymers in a finished oil.

Unlike mineral oils, which primarily exhibit Newtonian rheology, polymer-thickened oils exhibit viscoelastic rheology, that is, the viscosity is dependent on shear stress. When oils containing viscosity improvers are subjected to moderate shear stress, the viscosity decreases until it approaches the viscosity of polymer-free oil. When the stress is removed, the viscosity bounces back to the original viscosity level (Fig. 23). Known as temporary viscosity loss, this characteristic is desired in lubricants because it decreases the viscous drag at low temperatures and, hence, could contribute to fuel economy (Ref 33).

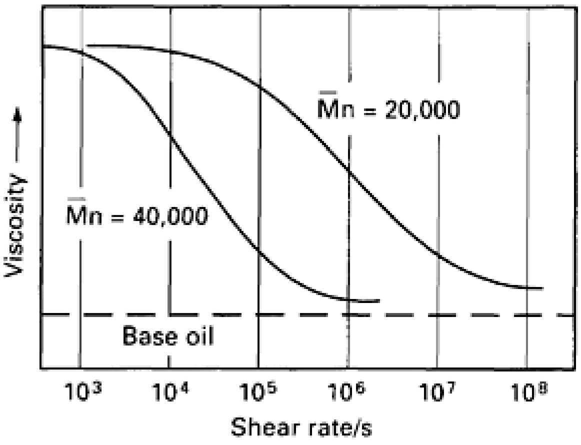


Fig. 23 Temporary shear loss. Source: Ref 32

When the viscosity-modified oil is subjected to a high shear stress, the polymer breaks down to lower molecular weight fragments, thereby causing a permanent viscosity loss, as shown in Fig. 24 (Ref 33). This type of loss is not desired because a formulated lubricant will not stay in its viscosity grade. The problem can be alleviated by preshearing the polymer prior to blending or by choosing a more shear-stable polymer.

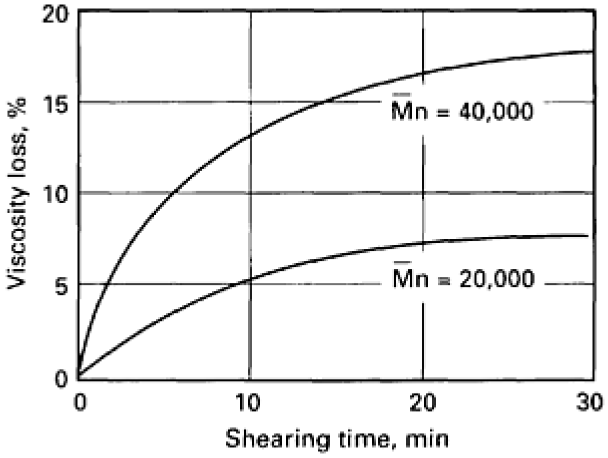


Fig. 24 Permanent shear loss. Source: Ref 32

In addition to mechanical breakdown (shear), polymers can undergo thermal and oxidative degradation. Thermally, polymers break down through chain scission and, ultimately, the polymer chains "unzip" back to the monomers. Oxidative degradation results from the breakdown of hydroperoxides, which are formed by the attack of oxygen on weak carbon-hydrogen bonds in the polymer molecule.

Viscosity improvers are primarily used in multigrade engine oils, gear oils, automatic transmission fluids, power steering fluids, greases, and some hydraulic fluids. Olefin copolymers are of the most popular type, followed by polymethacrylates, styrene-diene polymers, and styrene-polyesters.

Other Additives

In addition to the major classes of additives described above, lubricants contain a number of other additives. These include seal-swell agents, dyes, biocides, and couplers.

Seals are used in modern machinery for a variety of reasons. Their functions within lubrication systems are to:

- Isolate various lubrication environments from harmful elements
- Help maintain hydraulic pressure
- Allow removal and replacement of malfunctioning parts without necessitating the total dismantling of equipment
- Minimize contamination and the loss of lubricant

Seals are commonly made from polymeric materials, such as fluoro-elastomers, nitrile-rubber, polyacrylates, and silicones.

Certain base stocks and additive systems in lubricants can cause shrinkage, brittleness, and deterioration of the seals and impair the performance of the lubricating system. Seal-swell agents are additives that maintain the integrity of the seals. They include polyesters, some phosphorus derivatives, and a number of proprietary chemicals. These materials are commonly used in transmission and hydraulic fluids.

Dyes are used in lubricants as a color-code to ensure that the lubricant is used in the proper application and as a leak detection aid for the consumer. Automatic transmission fluids contain a red dye, whereas the two-stroke cycle oils contain a blue or purple dye. These dyes are oil-soluble organic compounds.

In general, mineral oil-based lubricants resist microbial attack because of their high-temperature operation and the presence of additives with biocidal action. However, high water-based lubricants, such as certain metalworking fluids and hydraulic fluids, are easily attacked by microbes. Control of bacterial growth is essential to minimize product deterioration and possible worker health hazards. This is accomplished by using water-soluble triazine, morpholine, and thiazoline derivatives, which possess biocidal properties. Triazines, which owe their biocidal action to their formaldehyde-releasing ability, find extensive use in this application.

Couplers are additives that are used in water-based lubricants to help stabilize microemulsions. Glycols are commonly used to fulfill this function.

Multifunctional Nature of Additives

There are a number of additives that perform more than one function. Zinc dithiophosphates, known mainly for their antiwear action, are also potent oxidation and corrosion inhibitors (Ref 25). Polyacrylates and styrene-polyesters can act as viscosity improvers, dispersants, and pour-point depressants. Basic sulfonates, in addition to acting as detergents, perform as rust and corrosion inhibitors, because they neutralize acidic products of fuel combustion and lubricant oxidation.

Lubricant Formulation

Lubricants are commonly formulated by using mineral oils and synthetic or partially synthetic base stocks (Ref 2). Recently, there has been a great deal of renewed interest in the use of base stocks of biological origin (vegetable oils and animal fats). Their biodegradable nature and nonpetroleum origin are two major reasons for this impetus.

Lubricant additives are usually supplied as performance packages that are blended in base stocks to yield formulated lubricants. These lubricants are expected to meet the performance requirements established by the SAE, API, and OEMs, as well as the end user. At high concentrations, the additives exist as association structures, with the polar ends being associated together. Upon dilution, de-aggregation occurs, and additive molecules attain a more active form.

Because the additives are reactive chemicals, they can interact with one another in the package, either synergistically or antagonistically (Ref 34). The job of the formulator is to deliver the intended performance by minimizing the antagonistic effects and by maximizing the synergistic effects of additives through careful balancing.

The performance package and the viscosity improver are usually sold separately. This is because the two are not always compatible when in concentrate form (Ref 31). When applications require a viscosity improver, it is blended in the base fluid, along with the performance package, to formulate the finished lubricant. Lubricant additive suppliers develop general-purpose performance packages that meet industry specifications using widely available base stocks. The supplier may tailor them for the individual company's use.

Tables 1 and 2 show the classes of additives used to formulate engine and nonengine lubricants for a variety of applications. All formulations do not contain all the additives identified in these figures.

Table 1 Engine lubricant formulations

Additive	Gasoline engine oils	Diesel engine oils	Natural gas engine oils	Aviation engine oils	Two-cycle engine oils
Dispersant	•	•	•	•	•
Detergent	•	•	•	•	•
Antiwear/EP agent	•	•	•	•	•
Oxidation inhibitor	•	•	•	•	•
Corrosion inhibitor/metal deactivator	•	•	•	•	•
Friction modifer	•	. . .	. . .	. . .	. . .
Pour-point depressant	•	•	•	•	. . .
Foam inhibitor	•	•	. . .	. . .	. . .
Viscosity improver	•	•	. . .	. . .	. . .
Other[(a)]	. . .	. . .	. . .	. . .	•

(a) Includes coupler, dye, diluent, and emulsifier

Table 2 Nonengine lubricant formulations

Additive	Automatic transmission fluids	Gear oils		Hydraulic fluids		Metalworking fluids	Greases
		Automotive	**Industrial**	**Tractor**	**Industrial**		
Dispersant	•	. . .	. . .	. . .	. . .	. . .	. . .
Detergent	•	. . .	. . .	•	•	•	
Antiwear/EP agent	•	•	•	•	•	•	•
Oxidation inhibitor	•	•	•	•	•	•	•
Corrosion inhibitor/metal deactivator	•	•	•	•	•	•	•
Friction modifier	•	•	•	•	. . .	•	. . .
Pour-point depressant	•	•	•	•	•	. . .	. . .
Foam inhibitor	•	•	•	•	•	•	. . .
Viscosity improver	•	•	•	•	•	. . .	•
Other[(a)]	•	. . .	. . .	•	•	•	•

(a) Includes antiseize, biocide, buffer, coupler, dye, emulsifer, seal-swell agent, and viscosity modifier

Introduction of a New Additive

The development of a new additive is initiated after a new product (lubricant) need is identified. The need for a new product, usually expressed by the OEMs and the end users, relates to either inadequate performance of existing products in current equipment or perceived needs of equipment under development. To fulfill this need, various organizations (such as SAE, API, ASTM, and OEMs) initiate the development of new performance specifications and the test methods.

The additive companies, either on their own or by collaborating with the lubricant supplier, try to satisfy the performance requirements established for the new product. If the additive company is unable to come up with the additive system based on their existing technology base, they initiate a project to develop and test a new additive.

The newly developed additives are blended with other additives in the base oil of a customer and are screened in a number of proprietary bench tests. The bench tests are accelerated tests that are devised to closely simulate the conditions the lubricant is expected to experience in actual service. This kind of testing is quite common, because it allows the evaluation of a large number of additives inexpensively.

Once the lubricant satisfies the performance criteria in the bench tests, full-fledged testing using actual equipment is carried out. This can be done in a laboratory or in collaboration with an end user. For additives used in automotive products, field trials may also be necessary. The costs associated with the development and testing of new additives can be phenomenal. The performance package that has successfully met all the performance requirements is ready to be marketed, either through factory-fill or the service-fill lubricant blenders.

References

1. F.X. Sieloff and J.L. Musser, "What Does the Engine Designer Need to Know About Engine Oils?", paper presented to Detroit Section of the Society of Automotive Engineers, Mar 1982
2. D.E. Ripple and J.F. Fuhrmann, "Performance Comparisons of Synthetic and Mineral Oil Crankcase Lubricant Base Stocks," paper presented at the Fourth International Colloquium (Esslingen, Germany), Jan 1984
3. K.L. Kreuz, Gasoline Engine Chemistry as Applied to Lubricant Problems, *Lubrication,* Vol 55, 1969, p 53-64
4. K.L. Kreuz, Diesel Engine Chemistry as Applied to Lubricant Problems, *Lubrication,* Vol 56, 1970, p 77-88
5. "Engine Service Classification System and Guide to Crankcase Oils Selection," No. 1509, American Petroleum Institute, 1985
6. "Engine Oil Viscosity Classification," J300, Society of Automotive Engineers, 1989 (revised)
7. "Engine Oil Performance and Engine Service Classification (Other than 'Energy Conserving')," J183, Society of Automotive Engineers, 1989 (revised)
8. H.E. Deen and J. Ryer, Automatic Transmission Fluids--Properties and Performance, *Fuels and Lubricants Technology: An Overview,* SP.603, Society of Automotive Engineers, 1984, p 117-127
9. R. Graham and W.R. Oviatt, "Automatic Transmission Fluids-Developments Toward Rationalization," paper presented at CEC International Symposium (Wolfsburg, Germany), June 1985
10. L.F. Schiemann and J.J. Schwind, Fundamentals of Automotive Gear Lubrication, *Fuels and Lubricants Technology: An Overview,* SP.603, Society of Automotive Engineers, 1984, p 107-115
11. "Axle and Manual Transmission Lubricant Viscosity Classification," J306, Society of Automotive Engineers, 1985 (revised)
12. "Lubricant Service Designations for Automotive Manual Transmissions and Axles," No. 1560, 5th ed., American Petroleum Institute, 1981
13. J.M. Sutherland, "Proposed Automotive Gear Lubricant Categories: Their Impact on the Industry," paper presented at National Petroleum Refiners Association National Fuels and Lubricants Meeting (Houston), 1989
14. "Ready Reference for Lubricant and Fuel Performance," No. 1288 240-94R, Lubrizol Corporation, 1988
15. T.C. Wilson, Modern Automotive Greases, *Fuels and Lubricants Technology: An Overview,* SP.603, Society of Automotive Engineers, 1984, p 101-106
16. A. Schilling, *Motor Oils and Engine Lubricants,* Scientific Publications, Great Britain, 1968

17. I.M. Feng, W.L. Perilstein, and M.R. Adams, Solid Film Deposition and Non-sacrificial Boundary Lubrication, *ASLE Trans.,* Vol 6, 1963, p 60-66
18. R.B. Jones and R.C. Coy, The Chemistry of Thermal Degradation of Zinc Dialkyldithiophosphate Additives, *ASLE Trans.,* Vol 24, 1981, p 91-97
19. A.D. Brazier and J.S. Elliot, The Thermal Stability of Zinc Dithiophosphates, *J. Inst. Petrol.,* Vol 53, 1967, p 63-76
20. N. Kharasch, Ed., *Organic Sulfur Compounds,* Vol 1, Pergamon Press, 1961
21. G.D. Thorn and R.A. Ludwig, *The Dithiocarbamates and Related Compounds,* Elsevier, 1962
22. W. Gerrard and H.R. Hudson, Chapter 13, *Organic Phosphorus Compounds,* Vol 5, G.M. Kosolopoff and L. Maier, Ed., Wiley Interscience, 1973
23. E. Cherbuliez, Chapter 15, *Organic Phosphorus Compounds,* Vol 6, G.M. Kosolopoff and L. Maier, Ed., Wiley Interscience, 1973
24. K.U. Ingold, Inhibition of Autoxidation of Organic Substances in Liquid Phase, *Chem. Rev.,* Vol 61, 1961, p 563-589
25. M.D. Johnson, S. Korcek, and M. Zinbo, Inhibition of Oxidation by ZDTP and Ashless Antioxidants in the Presence of Hydroperoxides at 160 °C, *Lubricant and Additive Effects on Engine Wear,* SP.558, Society of Automotive Engineers, 1983, p 71-81
26. H.H. Abou El Naga and A.E.M. Salem, Effect of Worn Metals on the Oxidation of Lubricating Oils, *Wear,* Vol 96, 1984, p 267-283
27. A.K. Vijh, Electrochemical Mechanisms of the Dissolution of Metals and the Contaminants Oxidation of Lubricating Oils Under High Temperature Friction Conditions, *Wear,* Vol 104, 1985, p 151-158
28. G. Scott, New Developments in the Mechanistic Understanding of Antioxidant Behavior, *J. Appl. Polym. Sci., Applied Polymer Symposium,* Vol 35, 1979, p 123-149
29. P. Becher, *Emulsions: Theory and Practice,* American Chemical Society Monograph Series, Reinhold, 1957, p 209-231
30. Th. F. Tadros, Ed., *Surfactants,* Academic Press, 1984, p 307
31. S.K. Baczek and W.B. Chamberlin, Petroleum Additives, *Encyclopedia of Polymer Science and Engineering,* Vol 11, 2nd ed., John Wiley and Sons, 1988
32. H.G. Muller, Mechanism of Action of Viscosity Index Improvers, *Tribol. Int.,* June 1978, p 189-192
33. R.W. Watson and T.F. McDonnell, Jr., Additives--The Right Stuff for Automotive Engine Oils, *Fuels and Lubricants Technology: An Overview,* SP.603, Society of Automotive Engineers, 1984, p 17-28
34. S.M. Hsu and A.L. Cummings, Interactions of Additives and Lubricating Base Stocks, *Lubricant and Additive Effects on Engine Wear,* SP.558, Society of Automotive Engineers, 1983, p 61-70

Solid Lubricants

Harold E. Sliney, National Aeronautics and Space Administration, Lewis Research Center

Introduction

SOLID LUBRICANTS are thin films composed of a single solid or a combination of solids introduced between two rubbing surfaces for the purpose of reducing friction and wear. Applications or operations involving severe temperatures, pressures, and environments, which preclude the use of organic fluids, have promoted the development of solid lubricants. Graphite and molybdenum disulfide (MoS_2) are the most frequently used inorganic solid lubricants. Graphite and molybdenite (the naturally occurring form of MoS_2) were used for lubrication purposes prior to the Industrial Revolution. In the early applications, molybdenite was probably used because it looks like graphite and was mistaken for it.

A significant amount of technical literature on the subject of solid lubricants is of relatively recent origin. Some sporadic technical references to solid lubricants appeared prior to 1940. For example, Paxton (Ref 1) cited an account by Salto in 1906 (Ref 2) that described the use of carbon as a brush material for electric motors, with an observation concerning the "unique self-lubricating properties of carbon." Winer (Ref 3) in his literature review of MoS_2 as a lubricant, cites some early (1939) patent literature that suggests the use of molybdenite as a lubricant; see for example, the work of Cooper and Damerell (Ref 4). The first technical journal article cited by Winer describes MoS_2 as a lubricant for vacuum (Ref 5). The earliest substantial body of technical literature on solid lubricants appeared in the middle and late 1940s. The solid lubricants research of that era to the present is well-documented. An extensive review of the subject is given in the chapter "Nonconventional Lubricants" in the book *Advanced Bearing Technology*, published by Bisson and Anderson in 1964 (Ref 6). More recent examples of reviews of solid lubrication were authored by Sliney (Ref 7) and Lancaster (Ref 8).

This article encompasses the period from about 1948 to the present (1991). Emphasis is placed largely, but not exclusively, on work performed at the NASA Lewis Research Center with a natural focus on aerospace applications. However, because of the generic nature of the research, the information presented in this article is applicable to most areas where solid lubricant technology is useful. The solid lubricants discussed include dichalcogenides (such as MoS_2), graphite, graphite fluoride, and polyimides. Less-conventional solid lubricants used under extreme temperature conditions, which include stable fluorides and lubricious oxides, are also described. Preparation techniques include air spraying, physical vapor deposition, plasma spraying, and powder metallurgy processes such as sintering and hot isostatic pressing.

Characteristic of Solid Lubricant Materials

There are some material properties that are characteristic of solid lubricants. Direct microscopic observations of the dynamics of solid lubrication show that sliding is accompanied by severe ductile shear of the solid lubricant film (Ref 9). This implies that to provide a low friction coefficient, the solid lubricant must have low shear strength. If the lubricant is crystalline, shear occurs by slip along preferred crystallographic planes. This slip is observed as a severe plastic flow in which the individual solid lubricant particles coalesce into a continuous, crystallographically oriented thin film that adheres to the lubricated surfaces and shears within the film. Low shear strength alone does not ensure lubrication if the material does not adhere to the lubricated surface. The importance of adherence of the solid lubricant film was emphasized by Peterson and Johnson (Ref 10) in a paper on the effect of crystal structure on the friction of possible solid lubricants.

Another property of importance is low abrasivity. This is a relative property that is a function of the hardness ratio of the lubricant and the bearing material. Obviously, the lubricant particles should be softer than the bearing material to avoid abrasive wear.

Finally, the solid lubricant must be thermodynamically stable in the environment of the application. A consideration of this property is important when selecting solid lubricants for high-temperature applications, but is equally important for moderate-temperature applications to ensure adequate storage stability and to avoid corrosion by atmospheric components such as oxygen and salt spray.

Layer Lattice Solid Lubricants

Layer lattice is a term used to describe crystal structures that consist of basal planes that are parallel to each other and consist of hexagonally oriented atoms (Fig. 1). The spacing between the planes is the *c*-spacing. The spacing between the atoms within the basal planes is the *a*-spacing. Compounds with a high c/a ratio have very an-isotropic shear properties with preferred shear parallel to the basal planes or perpendicular to the *c*-axis of the crystal structure. The dichalcogenides (disulfides, diselenides, and ditellurides) of molybdenum and tungsten have this structure. Brainard (Ref 11) showed that with the possible exception of tungsten diselenide, all of these compounds have lubricating properties, especially in vacuum. Table 1 lists these data giving the maximum temperature for lubrication with burnished dichalcogenide films in vacuum.

Table 1 Results of thermal stability and frictional experiments in vacuum of 10^{-7} to 10^{-4} MPa

Compound	Probable onset of thermal dissociation as detected by thermogravimetric analysis		Dissociation products first detected by mass spectrometry		Maximum temperature at which burnished films provided effective lubrication	
	°C	**°F**	**°C**	**°F**	**°C**	**°F**
MoS_2	930	1705	1090	1995	650	1200
WS_2	870	1600	1040	1905	730	1345
$MoSe_2$	760	1400	980	1795	760	1400
WS_2	700	1290	930	1705	760	1400
$MoTe_2$	700	1290	700	1290	540	1005

(a) Friction coefficient greater than 0.2 at all temperatures

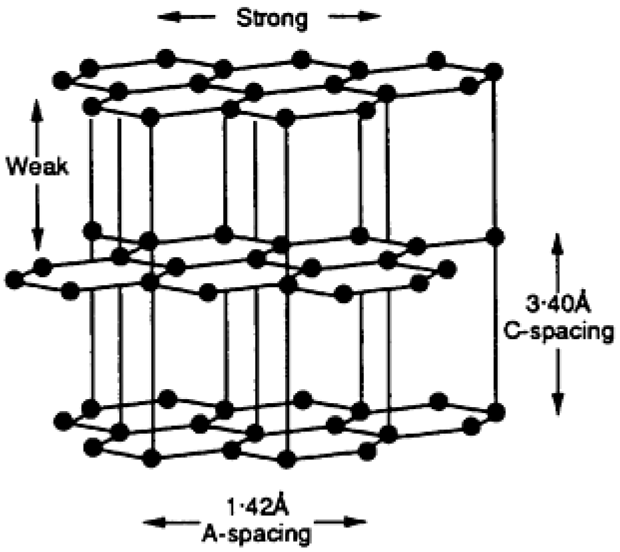


Fig. 1 Structure of graphite. The individual sheets consist of closely packed atoms, separated by a relatively large distance from neighboring sheets.

Compounds of this type have an AX_2 stoichiometry. Their crystal structures consist of hexagonal arrays of cations, each bonded to a pair of pendant anions (one on each side of the basal plane). Thus, a double layer of anions is present between each plane of cations. Strong, ionic bonding exists within each basal plane, but there are primarily weak Van der Waals forces within the double anion layers. Therefore, these compounds easily preferentially orient during sliding to allow slip of the basal planes nearly parallel to the sliding surfaces. This orientation behavior during sliding was demonstrated by electron and x-ray diffraction techniques early in the scientific investigation of MoS_2 (Ref 12, 13). Recent fundamental studies of sputtered MoS_2 films have shown that, even though reorientation occurs during sliding, the as-deposited crystal orientation has an influence on the wear life and chemical durability of the films (Ref 14, 15). Graphite also has a layered, hexagonal crystal structure, but does not appear to be an intrinsic solid lubricant. As will be discussed later, graphite requires adsorbed moisture or additions to develop lubricating ability.

It is also relevant that metals with a hexagonal crystal structure and a c/a ratio greater than about 1.6 exhibit predominantly basal plane slip, whereas those with a lower c/a ratio exhibit predominantly prismatic and pyramidal slip. Buckley (Ref 16) showed that those hexagonal metals that exhibit basal slip have lower friction coefficients in vacuum than the others.

MoS_2 and Other Dichalcogenides. The most widely used solid lubricant of this chemical class is MoS_2, the second is tungsten disulfide (WS_2), and the other dichalcogenides are found in scattered specialty applications. Both MoS_2 and WS_2 are intrinsic solid lubricants in that, unlike graphite, they do not require adsorbed materials or additives to develop lubricating capability. In fact, they lubricate best in the absence of such materials and are therefore lubricants of choice

for vacuum applications. Molybdenum disulfide is commonly used to lubricate moving mechanical components in space vehicles.

Molybdenum disulfide is applied to surfaces by a number of methods, including simple rubbing or burnishing, air spraying of resin-bonded or inorganically bonded coatings, and more recently, physical vapor deposition (PVD) techniques such as sputtering.

Burnished films are the easiest and least expensive to apply, but have very limited wear life. Resin-bonded spray coatings, especially the heat-cured variety, have superior wear life and are frequently used in ordinary air environments. They are typically 5 to 15 μm thick and have friction coefficients of about 0.06 to 0.15, depending on humidity and sliding conditions.

Early research with sputtered MoS_2 was reported by Spalvins and Przybyszewski (Ref 17). Their MoS_2 films, which they deposited in a direct current (dc) sputtering apparatus, were about 0.2 μm thick and were nearly stoichiometric. Figure 2 shows that the coatings had good frictional characteristics and good endurance in pin-on-disk tests in vacuum. However, the endurance life and friction coefficients of sputtered MoS_2 are not as good in air as they are in vacuum. The effect of atmospheric pressure on the friction of MoS_2-lubricated nickel is shown in Fig. 3; the friction coefficient increases from about 0.04 to 0.15 as the vacuum chamber pressure is increased above 13×10^3 Pa (100 torr).

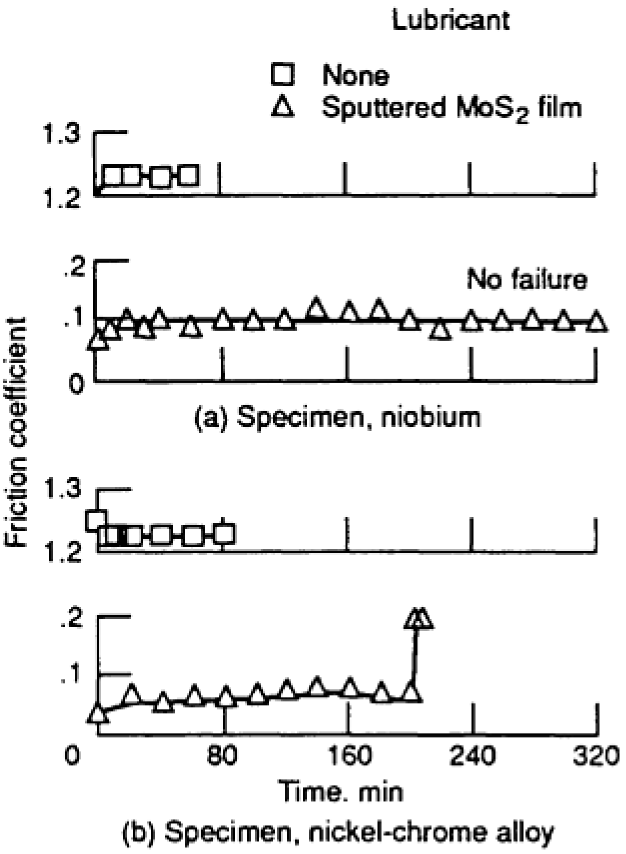


Fig. 2 Average friction coefficients of niobium sliding on two different specimens coated with sputtered MoS_2 in vacuum (10^{-11} torr, or 1.33×10^{-9} MPa). Load, 2 to 5 N (0.45 to 1.12 lbf); speed, 1.5 m/s (5 ft/min); ambient temperature

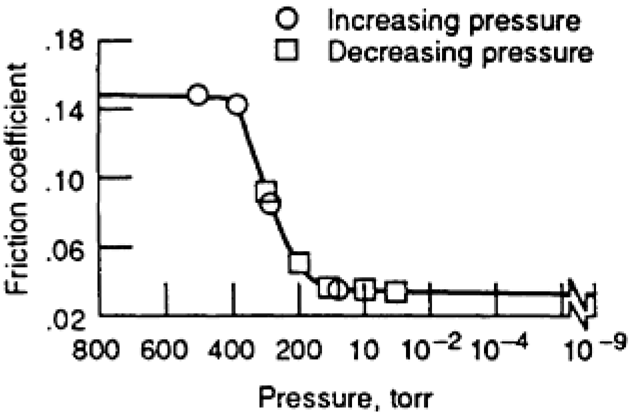


Fig. 3 Effect of pressure on coefficient of sliding friction for sputtered MoS^2. Load, 250 gf (0.55 lbf); speed, 40 rev/min; substrate/rider, Ni/Ni; room temperature. Source: Ref 18

Friction coefficients for sputtered MoS_2 as low as 0.01 in vacuum have been reported by Roberts (Ref 19). In the same reference, he also reports on the adverse effects of atmospheric contaminants. Oxygen and especially water vapor in room air causes a slow oxidative degradation of MoS_2, leading to early failure (Fig. 4). On the other hand, resin-bonded coatings of high quality have good shelf life and superior wear life in air, probably because of the oxidation protection afforded by the resin binder.

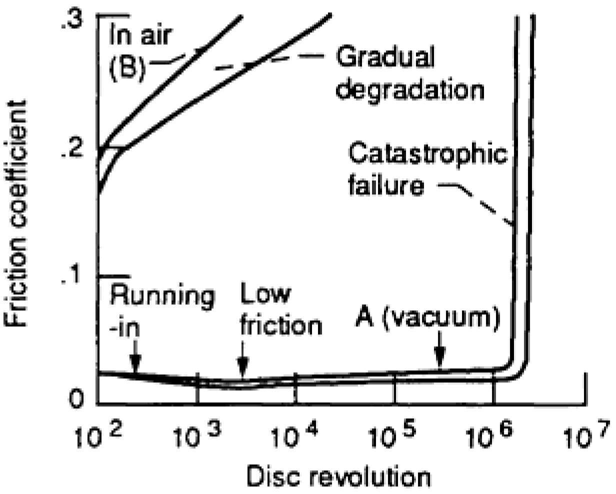


Fig. 4 Failure modes of sputtered MoS_2 under sliding motion. Source: Ref 19

Oxidation kinetic data for MoS_2 and WS_2, as determined by high-temperature x-ray diffraction, were reported by Sliney (Ref 20). Some oxidation kinetics data from this study for loosely compacted MoS_2 powders of 1 μm particle size are given in Fig. 5(a). The data show that the oxidation rate is strongly influenced by airflow rate through the reaction chamber. Increasing the airflow rate by a factor of about 6 increased the oxidation rate by a factor of 10 or more. At the lower flow of 5.5×10^{-6} mm^3/s (194×10^{-6} ft^3/s), MoS_2 was oxidized to molybdic oxide (MoO_3) in 1 h at 400 ° (750 °F). At the higher airflow rate, the temperature for an oxidation half-life of 1 h was reduced to 300 °C (570 °F). Figure 5(b) compares the oxidation kinetics of MoS_2 and WS_2 at the lower airflow rate. The curves intersect with MoS_2 oxidizing more rapidly than WS_2 above 340 °C (645 °F). Comparison of these data with the results of friction experiments shows that the loss of lubricating ability of these compounds in air coincides with the temperatures at which rapid oxidation occurs (Fig. 6).

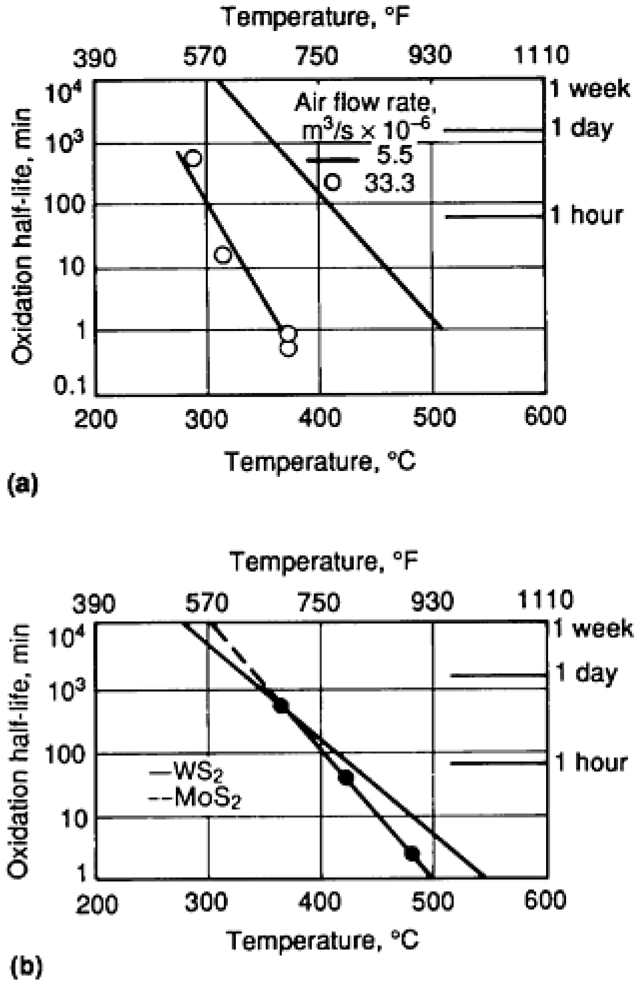


Fig. 5 Oxidation kinetics of MoS_2 and WS_2 with average particle size of 1 μm and compact density of 50%. (a) Oxidation characteristics of MoS_2 at two airflow rates. (b) Comparative oxidation of WS_2 and MoS_2 with an airflow rate of 5.5×10^{-6} m^3/s. Source: Ref 20

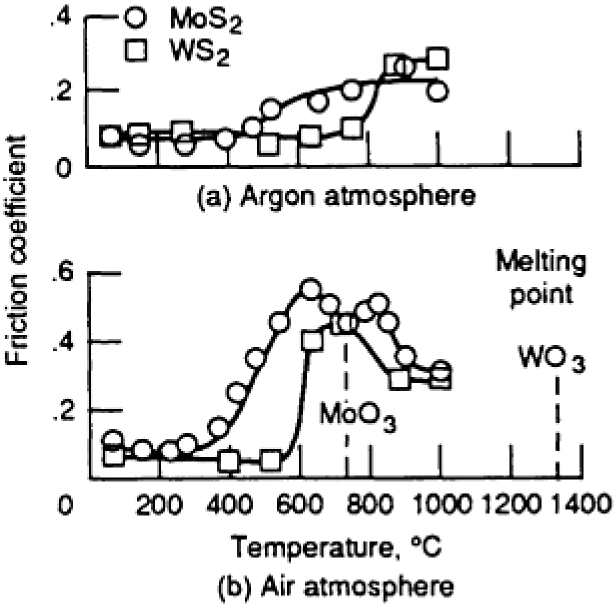


Fig. 6 Variation of friction coefficient with temperature and MoS_2 and WS_2 in (a) argon and (b) air in pin-on-

disk tests using steel pins with 4.576 mm (0.183 in.) hemispherical tip radius. Force, 9.6 N (2.2 lbf); speed, 0.025 m/s (0.082 ft/s)

Graphite. The largest-volume tribological uses of carbon are as electrically conducting brushes for motors and generators and as the rubbing element in mechanical seals. The graphite form of carbon, which has a hexagonal, layered crystal structure (Fig. 1), also has application as a powder or bonded solid-lubricant coating. In spite of its desirable crystal structure, graphite is not an intrinsic solid lubricant. It lubricates in a normal air atmosphere, but fails to lubricate at high altitudes or in vacuum. Savage (Ref 21) reported convincing experimental evidence that graphite must adsorb moisture or some other condensable vapor such as hydrocarbons in order to be lubricative. This model appeared to be substantiated by the observation that, even in the presence of condensable vapors, graphite failed to lubricate above the desorption temperature.

However, Peterson and Johnson (Ref 10) observed that graphite again becomes lubricative at high temperatures when the lubricated metal becomes visibly oxidized. Some of their results showing the effect of temperature on the friction coefficient of graphite-lubricated cast Inconel sliding on Inconel X-750 in air are shown in Fig. 7. The authors postulate that restoration of lubrication at about 425 °C (800 °F) was due to interaction of graphite with oxides of the metal. The surface oxides were thought to promote adhesion of the graphite to the lubricated surfaces. The maximum temperature for lubrication with graphite films is limited by oxidation to about 550 °C (1020 °F).

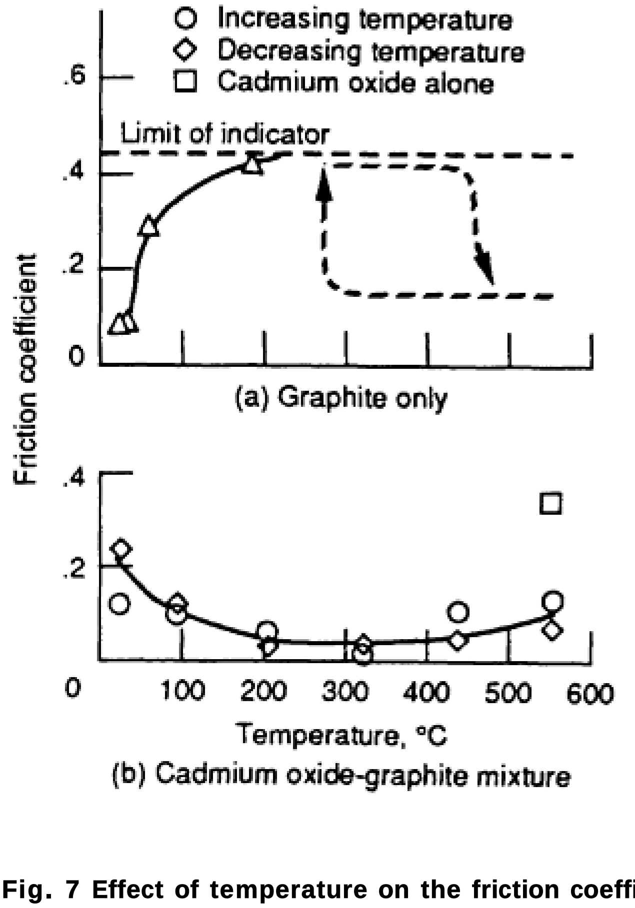


Fig. 7 Effect of temperature on the friction coefficient of cast Inconel sliding on Inconel X-750 in air lubricated by (a) graphite and (b) cadmium oxide-graphite. Source: Ref 10

Physical Vapor Deposition (PVD) of Tribological Coatings. Vapor-deposited coatings for tribological applications are of two major types: soft solid-lubricant coatings, and hard wear-resistant coatings. The basic processes for applying PVD coatings are simple vapor deposition, sputtering, and ion plating (see the article "PVD and CVD Coatings" in this Volume). In recent years various ion beam assisted deposition methods, which are modifications of the more basic process, have been used to deposit very adherent tribological coatings (Ref 22, 23). Sputtered MoS_2 was discussed earlier in the section on dichalcogenide solid lubricants. Two other categories of PVD coatings that are tribologically significant are sputtered hard coatings for wear control and ion-plated soft metals for lubrication.

Sputtered Hard Coatings. The hardness and oxidation characteristics of some hard coating materials are given in Table 2. These very hard materials are brittle. Therefore, they must have very good adhesion to the substrate to avoid delamination, and they must not be too thick if they are to withstand bearing surface stresses and substrate elastic deformations under dynamic (sliding, rolling) contact conditions. Acceptable coating thicknesses are typically 0.2 to 1.0 μm. Hard coatings alone can provide wear resistance. They are also used as an intermediate coating between the substrate and a solid-lubricant film.

Table 2 Bulk properties of some hard coat materials

Data from *Engineering Properties of Ceramic Materials,* Battelle Memorial Institute, published by American Ceramic Society, Columbus, OH, 1966

Material	Microhardness, kg/mm^2	Oxidation temperature[(a)]	
		°C	**°F**
B_4C	4200	1090	1995
TiC	3200	540	1005
SiC	2900	1650	3000
Cr_3C_4	2650	1370	2500
WC	2050	540	1005
Si_3N_4	2000	1400	2550
TiN	1950	540	1005

(a) Temperature for appreciable detrimental oxidation (passivating oxide films form at lower temperatures).

(b) Estimated conversion from published Moh hardness of 9

Figure 8 compares the endurance lives in vacuum of MoS_2 lubricated ball bearings with and without an intermediate 0.1 μm hard coating of sputtered Cr_3Si_2 on the raceways. Sputtered 0.2 μm thick MoS_2 coatings were applied to the raceways and the cages of the bearings. The bearings without the intermediate coating had reasonably good life of about 200 h, but those with the hard coating did not fail in 1000 h, at which time the tests were stopped. Figure 9 shows that 1000+ h bearing life can also be achieved without the intermediate hard coating by sputter coating the bearing balls, in addition to the cage and raceways, with MoS_2. The combined effect of MoS_2 on all surfaces and an intermediate hard coating on the raceways and/or the balls was not investigated.

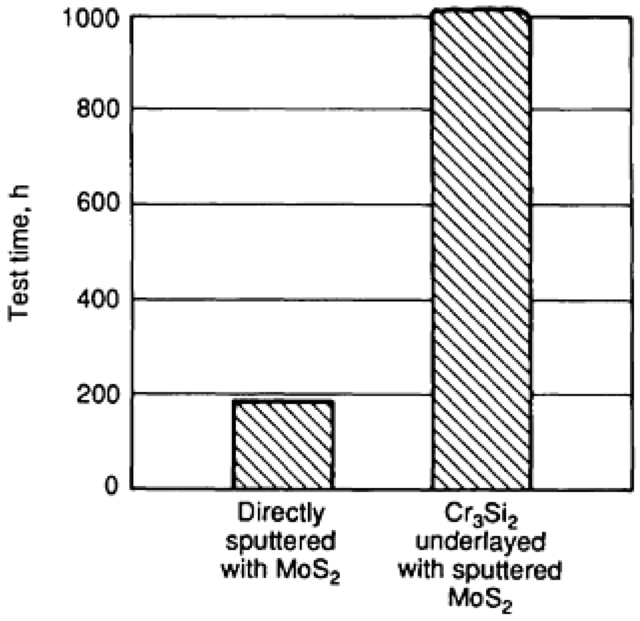


Fig. 8 Endurance lives of type 440C stainless steel ball bearings with sputtered MoS_2 films on races and cage--with and without a Cr_3Si_2 underlayer. Source: Ref 24

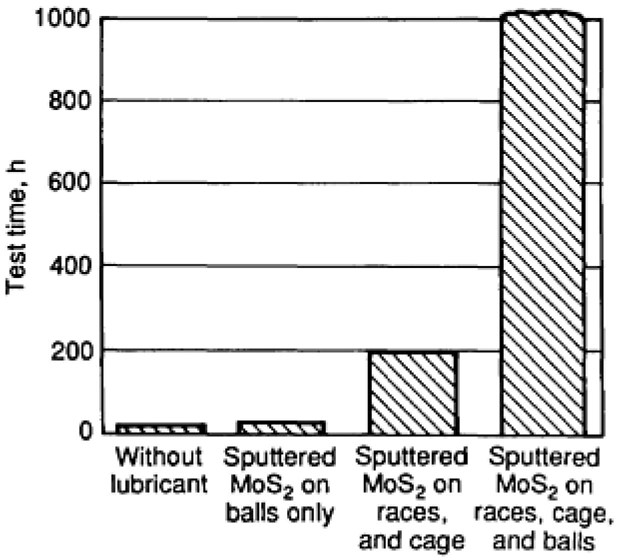


Fig. 9 Endurance lives of type 440C stainless steel ball bearings with and without sputtered MoS_2 film. Source: Ref 24

Graphite fluoride (CF_x) is not as well known as MoS_2 and graphite, but it has some unique properties that make it attractive as a solid lubricant. It is an electrical insulator (unlike graphite) and is nonwettable by water. The frictional properties are less influenced by humidity than either MoS_2 or graphite. Graphite fluoride is also known as carbon monofluoride because of its nominal 1:1 carbon-to-fluorine stoichiometry. However, the stoichiometry can be controlled from about x = 0.25 to 1.1. For x = 1.0 to 1.1, the compound is pure white and is preferred for applications were discoloration by the lubricant must be avoided.

Wear life and friction coefficient data from pin-on-disk tests reported by Fusaro and Sliney (Ref 25) for burnished films of CF_x on sand-blasted type 440C steel are shown in Fig. 10. The corresponding data for MoS_2 are included as a baseline for comparison. Film failure was taken as the time at which the friction coefficient reached 0.3. A very dry air atmosphere was used because of the deleterious effect of moisture on lubrication with burnished MoS_2. The CF_x films were the more durable over the entire temperature range. Friction coefficients were well below 0.1 for both compounds up to their failure time or temperature. Failure temperatures of 400 °C (750 °F) for MoS_2 and 480 °C (895 °F) for CF_x correlate well with the oxidation kinetics results for MoS_2 shown in Fig. 5 and the thermal decomposition temperature for CF_x reported by Kuriakose and Margrave (Ref 26). Therefore, CF_x was the better lubricant under the conditions of these tests. McConnell *et al.* (Ref 27) reported that MoS_2 had superior load-bearing capacity in tests on a Falex test machine in which double V-blocks are loaded against a rotating cylinder coated with the test lubricant. However, the load-carrying capacity of CF_x is considered by this author to be adequate for most conforming sliding contacts, especially when combined with a suitable binder such as a polyimide varnish. Fusaro (Ref 28) in his paper on the mechanism of graphite fluoride lubrication describes the lubricating process with this material as consisting of plastic flow of thin films of graphite fluoride between the sliding surfaces.

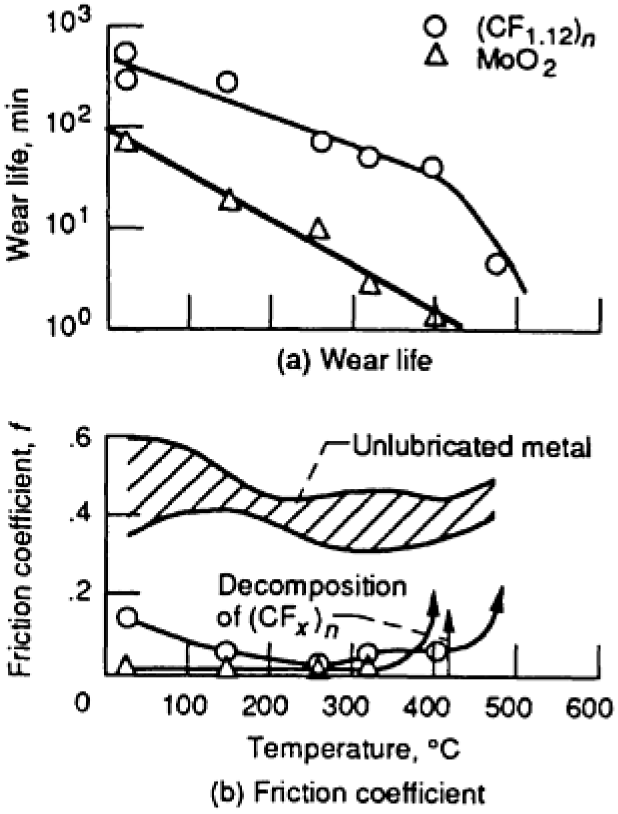


Fig. 10 Effect of temperature on wear life and friction coefficient of graphite fluoride [$(CF_{1.12})_n$] and MoS_2 powders burnished on sand-blasted type 440C stainless steel disks. Riders, 440C stainless steel with 4.76 mm (0.19 in.) hemispherical tip radius; linear sliding speed, 1.6 m/s (5.2 ft/s); load, 5 N (1.1 lbf); atmosphere, dry air (20 ppm H_2O). Source: Ref 25

Polyimide and Polyimide-Bonded CF_x Coatings. Data reported by Fusaro (Ref 29) from pin-on-disk experiments with polyimide varnish coatings on type 440C steel are given on Fig. 11, where the endurance lives and friction coefficients are shown as a function of temperature. Results are compared for three different controlled atmospheres: dry argon, dry air, and air containing 10^4 ppm water vapor. The data show that there is a sharp transition in the life and friction between 25 and 100 °C (75 and 212 °F), above which lower friction and much longer wear life were observed. This transition in the friction and endurance of polyimide coatings has been attributed to a second order relaxation in the molecular bonds of the polymer (Ref 30).

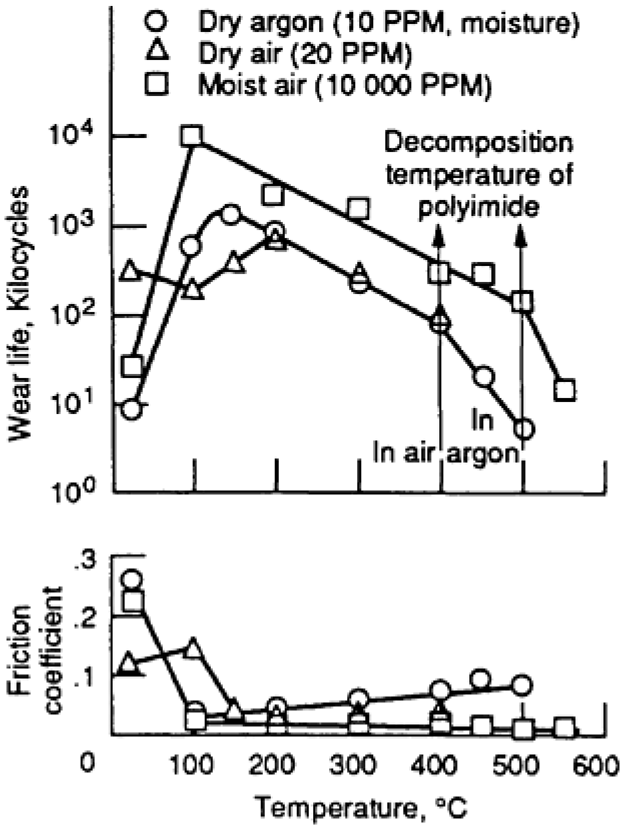


Fig. 11 Friction coefficient and wear life as a function of temperature for polyimide thin films in various atmospheres. Pin-on-disk tests run with type 440C stainless steel pins with 4.76 mm (0.19 in.) hemispherical tip radius. Load, 9.6 N (2.2 lbf); speed, 2.7 m/s (8.9 ft/s). Source: Ref 29

The effect of adding graphite fluoride or MoS_2 to the polyimide varnish is shown in Fig. 12. The coating formulations contained three parts by weight MoS_2 to one part polyimide solids dissolved in a solvent, or three parts of CF_x to two parts of polyimide solids. The density of CF_x is about one-half that of MoS_2; therefore, the two formulations contained equal volume percentages of solid lubricant. The solid lubricant additions clearly reduce the friction and greatly increase the wear life of the coatings at room temperature. The effect of the molecular relaxation transition on the tribological properties is completely eliminated by the solid-lubricant additions. The CF_x addition also resulted in improved coating endurance at all temperatures to 500 °C (930 °F), whereas MoS_2 addition deteriorated the endurance of the coating relative to the base polyimide films.

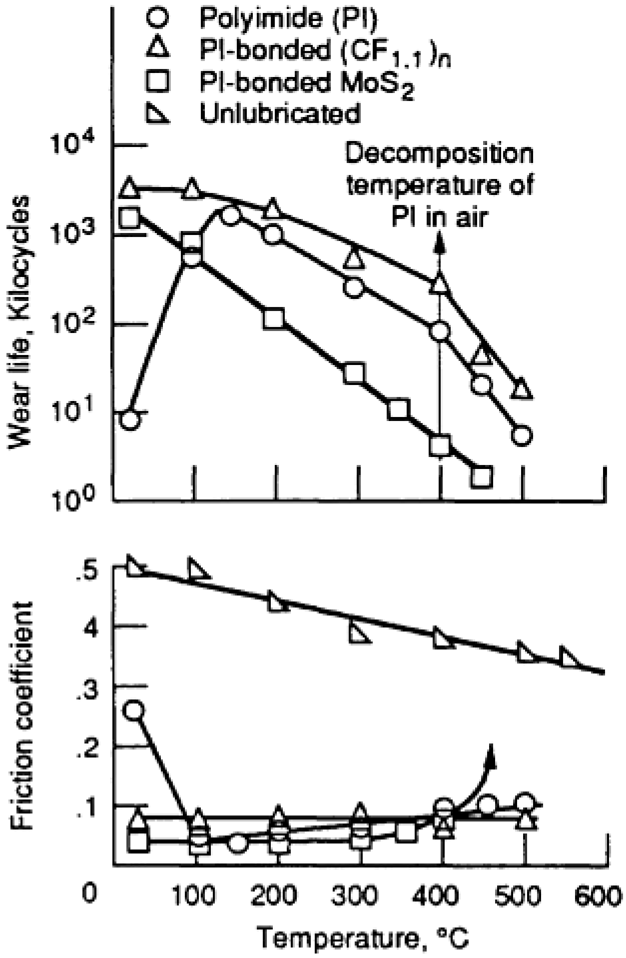


Fig. 12 Friction coefficient and wear life as a function of temperature for three solid lubricant films run in dry air and moist air (10 × 10^4 ppm H_2O) against type 440C stainless steel pins with 4.76 mm (0.19 in.) hemispherical tip radius. Load, 9.6 N (2.2 lbf); speed, 2.7 m/s (8.9 ft/s). Source: Ref 31

Wagner and Sliney (Ref 32) have reported on polyimide-bonded CF_x films that are excellent back-up lubricants for compliant (foil) gas bearings at temperatures up to about 350 °C (660 °F). This is about 100 °C (180 °F) higher than the limit for PTFE films that are commonly used in foil bearing applications.

Fusaro (Ref 33) has published a comprehensive review of the tribological properties of polyimide coatings and polyimide-base self-lubricating composites.

Polymer Composites. Many polymers, including nylon, acetal, polyethylene, and polytetrafluoroethylene (PTFE), are used in the unfilled or filled forms as bearing materials. Solid-lubricant fillers are typically powdered graphite, MoS_2, or CF_x. The powdered additives can improve the friction and wear properties, but usually weaken the material and, as a consequence, reduce their load-carrying capability. Where maximum load capacity is needed, fiber reinforcement is used. The fibers are in the form of woven fabric, chopped fibers, or filament windings. Glass fabric is often used to reduce creep of bonded PTFE liners in airframe bearings and other heavily loaded bearings. Polyester fibers have been found to be an effective nonabrasive reinforcement for PTFE bearing liners.

Sliney and Williams (Ref 34) reported on tests of nonmetallic plain cylindrical bearings consisting of an aramid (Kevlar) fiber-reinforced shell with a bonded, self-lubricating woven liner of polyester fiber-reinforced PTFE. The results are summarized in Fig. 13 and show that low friction and wear and load capacities in excess of 207 MPa (30 ksi) are achieved from -23 to 121 °C (-9 to 250 °F). The bearings are limited by creep deformation at higher temperatures.

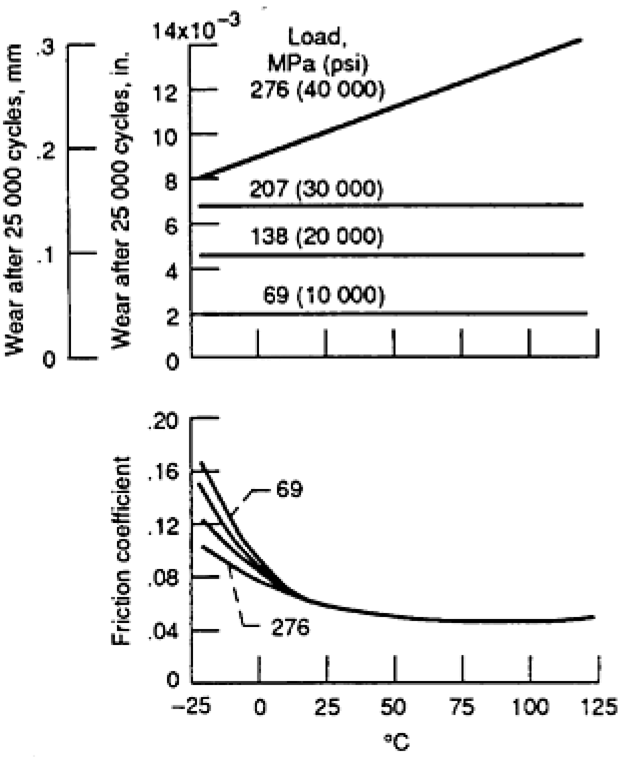


Fig. 13 Effect of temperature on wear and friction of PTFE-lined composite bushings. Bore diameter, 0.025 m (0.082 ft); journal oscillation at ±25° and 10 cycles/min. Source: Ref 34

Considerable research has been done on graphite fiber-reinforced polymers as self-lubricating bearing materials. Giltrow and Lancaster (Ref 35, 36) reported on the friction and wear properties of composites incorporating chopped graphite fibers in thermoplastic and thermosetting resins including high-temperature polyimide. In addition to high strength, graphite fibers have the additional advantages of high thermal conductivity and lubricating ability. Graphite fiber-reinforced polyimide (GFRPI) is of particular interest for aeronautics applications and has been extensively researched for airframe and gas turbine engine applications. The GFRPI composites investigated at the NASA Lewis facility contained chopped graphite fibers and were prepared by transfer molding and curing under heat and pressure.

The results of pin-on-disk bench tests for type 440C steel pins sliding against GFRPI were reported by Fusaro and Sliney (Ref 37). Favorable tribological properties were observed with several combinations of addition (A)-type and condensation (C)-type polyimides and low-modulus (L-type) and high-modulus (H-type) fibers. Figure 14 gives the friction coefficients and wear track profiles of the various composites at room temperature and at 300 °C (570 °F). In these tests, the best results were obtained with an addition-type polyimide reinforced with low-modulus graphite fibers (AL composite).

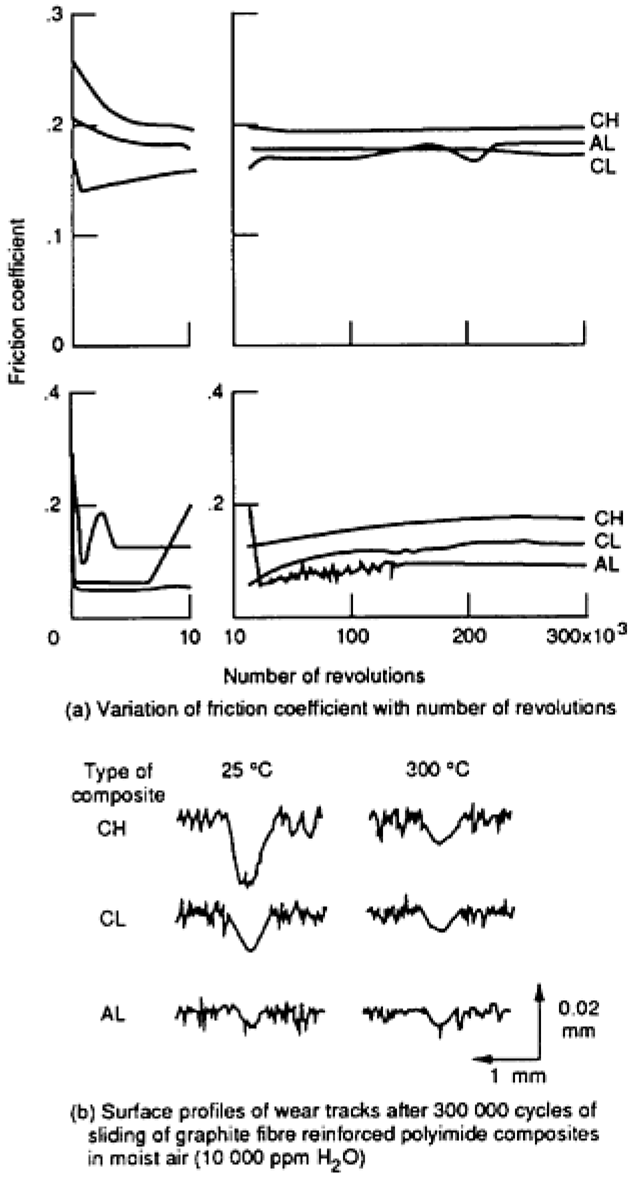


Fig. 14 Tribological performance of polyimide composite disks sliding against type 440C stainless steel pins with 4.76 mm (0.19 in.) hemispherical tip radius. Load, 9.6 N (2.2 lbf); speed, 2.5 m/s (8.2 ft/s). Source: Ref 37

Bearing Tests of GFRPI. Plain cylindrical and plain spherical bearing designs with self-lubricating GFRPI surfaces were tested to 350 °C (660 °F) in several test programs at NASA Lewis. The spherical bearing designs evaluated are shown in Fig. 15. The first bearings tested had molded spherical elements of AL-type GFRPI with various graphite contents in steel outer rings. Friction coefficients for various fiber loadings are shown in Fig. 16 for bearing temperatures from 25 to 350 °C (75 to 660 °F). The results for a conventional glass fiber reinforced bearing are included for comparison. The composite with the highest graphite fiber content provided the lowest bearing torque but failed by brittle fracture at 315 °C (600 °F). The next lowest bearing torque was provided by the composite with 45 wt% of graphite fibers and no failures occurred with a 35 MPa (5000 psi) radial load. A 1:1 fiber-to-polymer ratio was used in all subsequent tests. The standard PTFE-lined bearing had very low torque to 200 °C (390 °F), but failed when the liner extruded out of the bearing at 250 °C (480 °F).

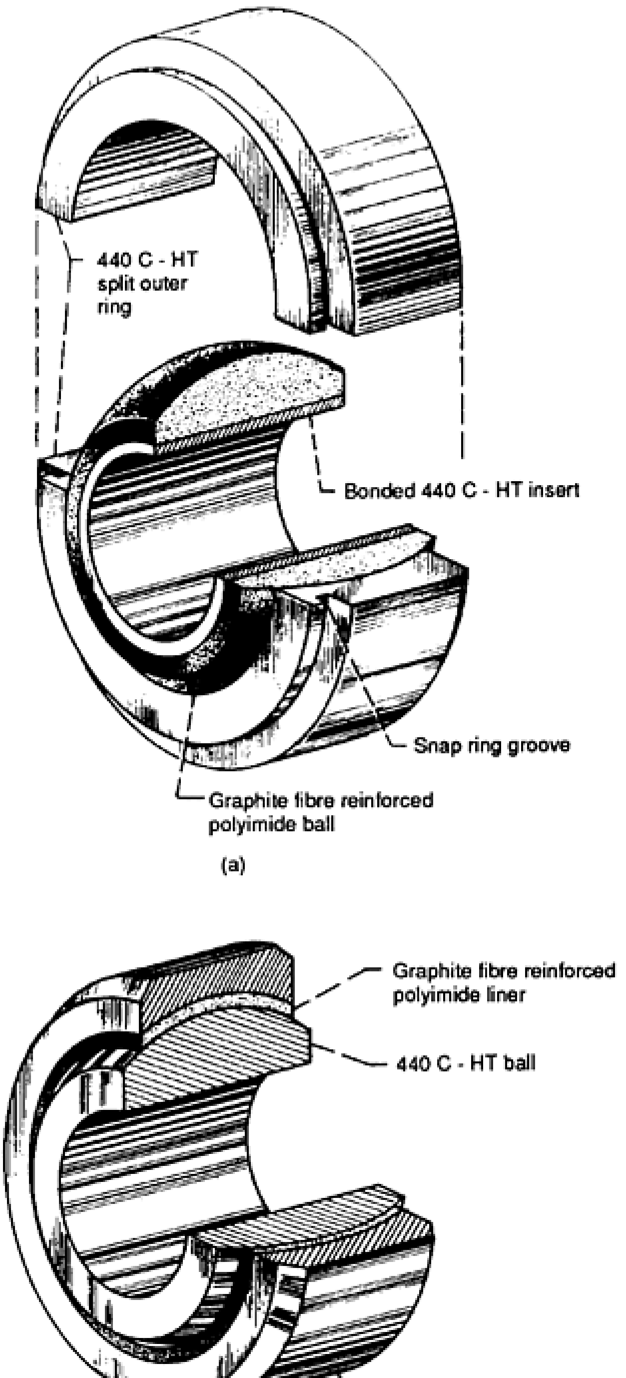


Fig. 15 Test bearings employing graphite fiber reinforced polyimide (GFRPI). (a) Bearing design featuring GFRPI ball. (b) Bearing design featuring a steel ball and a GFRPI liner bonded to a steel outer ring

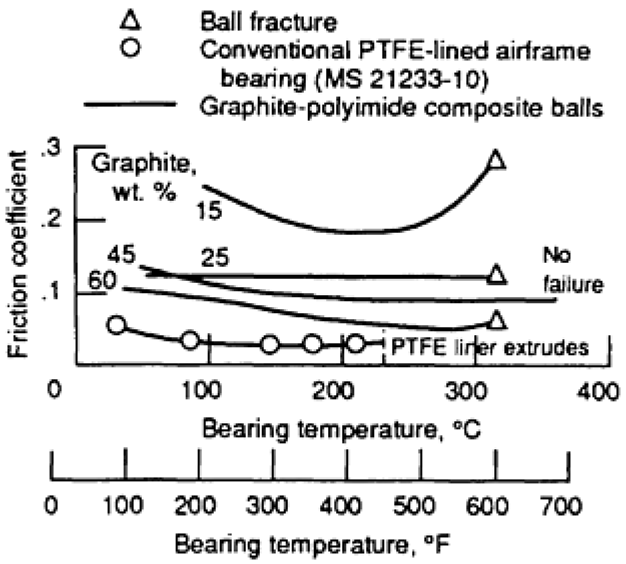


Fig. 16 Friction of spherical bearings with polyimide-graphite fiber composites of various fiber contents. Stellite 6B journal. Radial unit load, 35 MPa (5000 psi); journal oscillation in cylindrical bore at 1 Hz, ±15°

Several modifications of the bearing designs shown in Fig. 15 were compared (Ref 39). Figure 17 shows that no significant difference was noted in bearing friction for the two bearing designs. However, a much higher load capacity is achieved with the design shown in Fig. 15(b), which consists of a steel ball and a thin (1.5 mm, or 0.06 in.) GFRPI liner bonded to a steel outer ring (Ref 40). The dynamic load capacities of cylindrical bushings with GFRPI liners had about the same dynamic load capacities as the GFRPI-lined sphericals if a liner retention step was provided at each end of the bushing (Ref 41).

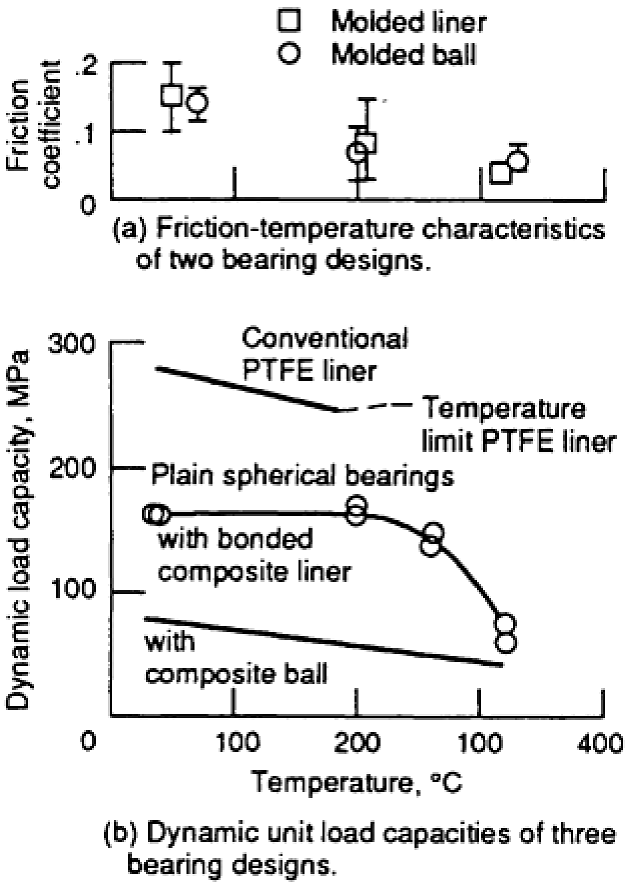


Fig. 17 Friction and dynamic load capacities of bearings lubricated with GFRPI composite. Source: Ref 39

Gardos and McConnell (Ref 42) have published an extensive four-part report on the preparation and performance of self-lubricating polyimide composites reinforced with three-dimensional woven graphite fibers. The purpose of the three-dimensional weave is to afford maximum strength and load capacity. Load capacities up to 200 MPa (29 ksi) were obtained during low-speed oscillating tests.

Extreme-Temperature Solid Lubricants

All of the solid lubricants so far discussed suffer thermal decomposition or oxidative degradation at or below 550 °C (1020 °F). For higher temperatures, classes of materials that have shown some promise are lubricious oxides and some fluorides of alkali metals, alkaline earth metals, and rare earth metals.

The tenacious oxides formed on nickel-base and cobalt-base superalloys at high temperatures have a profound effect on reducing friction and wear (Ref 43). Lankford *et al.* (Ref 44) showed that certain metals, when ion beam mixed into the surface of zirconia and silicon nitride ceramics, reduce their friction against some counterface materials at 800 °C (1470 °F). The beneficial effect is attributed to oxidized films of the implanted metals.

Oxidatively stable fluorides, such as CaF_2 and BaF_2, are lubricious from about 400 to 900 °C (750 to 1650 °F) (Ref 45). Other fluorides that have shown some lubricating capability at high temperatures are LiF, MgF_2, and fluorides of some rare earth metals, particularly, LaF_3 and CeF_3 (Ref 46, 47). Fluorides have been applied as (1) ceramic-bonded coatings using a porcelain enameling process, (2) fusion-bonded all-fluoride coatings, and (3) constituents of plasma-sprayed composite coatings. Most recently, they have been incorporated as components of free-standing powder metallurgy composites. The results of this research are summarized in the following section.

Ceramic-Bonded Fluorides. Early research on ceramic coatings and ceramic-bonded CaF_2 coatings at NASA Lewis was summarized by Johnson and Sliney (Ref 43) and Sliney (Ref 48). Detailed references are given in the bibliographies of those papers. The coatings are prepared by spraying and firing procedures similar to those used in porcelain enameling,

but the compositions are very different compared to conventional porcelain enamels and are specially formulated with the objective of achieving good tribological properties. Because at least some of the components of these coatings melt during the firing process, they are referred to as "fusion-bonded" coatings. In most cases the coating thickness is 0.02 mm, ±0.01 mm (0.0008 in., ±0.0004 in.).

In a formulation study of ceramic compositions, a CoO-base coating and a BaO-base coating for nickel-chromium superalloys were developed, both of which provided tribological benefit to at least 800 °C (1500 °F). High surface temperatures due to frictional heating at high sliding velocities or high bulk specimen temperatures were beneficial to lubrication with these coatings. The friction and wear for cast Inconel sliding on uncoated and on coated Inconel X-750 at a sliding velocity of 2.2 m/s (7.2 ft/s) are shown in Fig. 18. Both coatings provided a considerable reduction in pin wear compared to the unlubricated specimen. Friction was reduced over the entire temperature range with CoO-base coating and up to about 600 °C (1110 °F) with the BaO-base coating. Lower friction was obtained when the CoO-base composition was used as a binder for CaF_2. Friction and wear for coatings prepared from blended powders with a 3:1 CaF_2-to-binder ratio (where the binder is CoO-base ceramic) are shown in Fig. 19. Friction coefficients were about 0.25 at room temperature and gradually decreased to about 0.15 at 800 °C (1470 °F). Very careful coating procedures are necessary to achieve a uniform distribution of CaF_2 throughout the ceramic binder. The CoO-base ceramic is highly vitrified (glassy) and is temperature-limited to the maximum temperature at which the glass phase will not act as an adhesive upon subsequent cooling below its softening point.

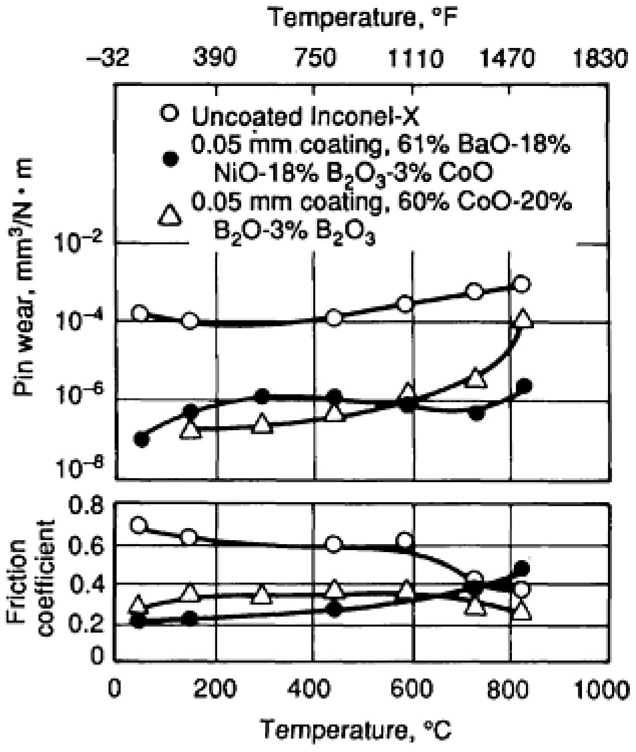


Fig. 18 Effect of temperature on lubricating properties of CoO-base and BaO-base coatings on Inconel X-750 sliding against Inconel X-750 pins with a 4.76 mm (0.19 in.) hemispherical tip radius. Load, 9.6 N (2.2 lbf); speed, 2.2 m/s (7.2 ft/s)

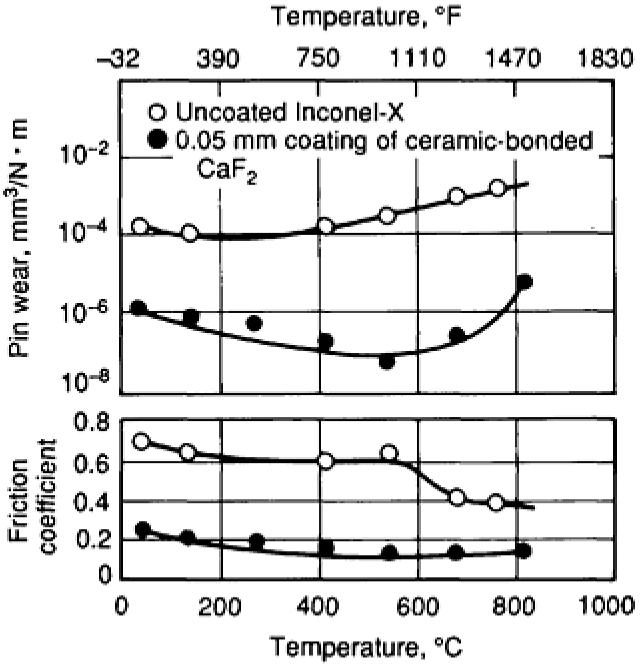


Fig. 19 Effect of temperature on lubricating properties of ceramic-bonded CaF_2 coatings on Inconel X-750 sliding against Inconel X-750 pins with a 4.76 mm (0.19 in.) hemispherical tip radius. Load, 9.6 N (2.2 lbf); speed, 2.2 m/s (7.2 ft/s)

Another approach involved preparing coatings from composite powders to avoid segregation during spraying and firing (fusion bonding) of the coatings. A composition near the melting point for the CaF_2-Ca_2SiO_4-CaO ternary system was chosen, and the CaF_2-BaF_2 binary eutectic was substituted for CaF_2 in order to lower the melting point. The coating was fusion bonded by firing in air at 980 °C (1795 °F). The friction coefficient as a function of temperature at two sliding velocities is shown in Fig. 20. Friction was quite high at the lower sliding velocity of 2.5 m/s (8.2 ft/s); the friction coefficient was 0.5 at room temperature and gradually decreased to about 0.3 at 600 to 800 °C (1110 to 1470 °F). However, at 10 m/s (33 ft/s), friction coefficients were less than 0.2 at all temperatures. Even at the lower speed, the pin wear factors were 1/1000th those of the uncoated metal, or about 10^{-6} $mm^3/N \cdot m$.

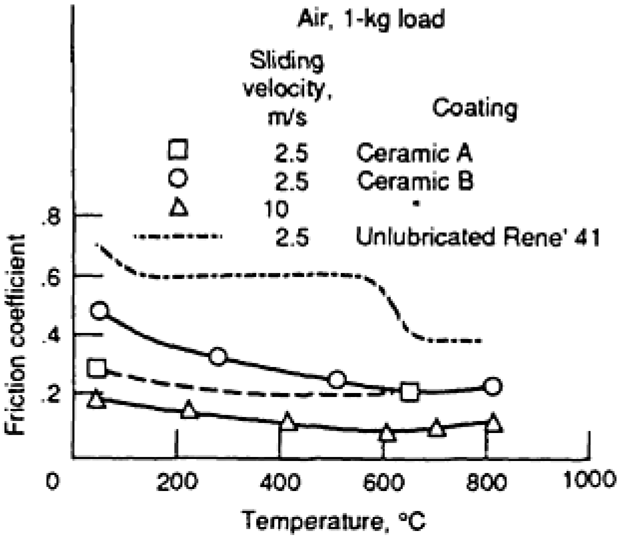


Fig. 20 Friction-temperature characteristics of fluoride coatings modified with silicate and oxide additions. Ceramic A: 60.8CaF_2-13.1LiF-17.4Ca_2SiO_4-8.7CaO. Ceramic B: 48.3 BaF_2-21.7CaF_2-20Ca_2SiO_4-10CaO

Fused Fluoride Coatings. Sliney *et al.* (Ref 45) reported on the performance of coatings consisting entirely of stable fluorides with emphasis on compositions from the BaF_2-CaF_2 binary system. Results of friction and wear tests are shown in Fig. 21. Over the entire temperature range of 25 to 816 °C (75 to 1500 °F) and at two sliding velocities of 2.3 and 10 m/s (7.5 and 33 ft/s), the coatings were responsible for very substantial reductions in friction and wear relative to the uncoated metals. At 2.2 m/s (7.2 ft/s), the friction coefficient at room temperature was rather high at 0.4 and decreased with temperature to 0.2 at 260 °C (500 °F) and 0.11 at 800 °C (1470 °F). At 10 m/s (33 ft/s), friction coefficients were 0.2 at room temperature, gradually decreasing to about 0.15 at 800 °C (1470 °F). No coating failures occurred in one hour tests--52.2 × 10^3 cycles at 2.3 m/s (7.5 ft/s) and 227 × 10^3 cycles at 10 m/s (33 ft/s).

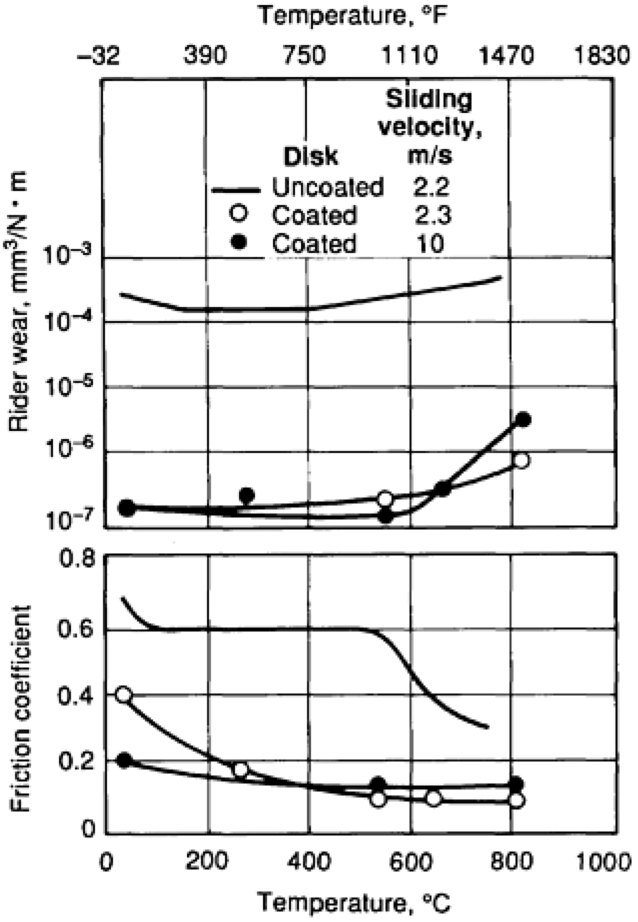


Fig. 21 Lubricating properties of fused fluoride coating composition in air at two sliding velocities. Coating composition, $38CaF_2$-$62BaF_2$; coating thickness, 0.025 to 0.050 mm (0.001 to 0.002 in.); duration of run, 1 h; load, 500 gf (1.1 lbf); specimens, pins with 4.76 mm (0.19 in.) hemispherical tip radius sliding against coated disks. Source: Ref 45

The relatively high friction at low temperature and sliding velocities led to an investigation in which it was found that the addition of silver to the BaF_2-CaF_2 coatings could result in a uniformly low friction from room temperature to 800 °C (1470 °F). This is illustrated in Fig. 22, which compares the friction-temperature characteristics of fluoride coatings with and without a 35 wt% silver addition (Ref 49).

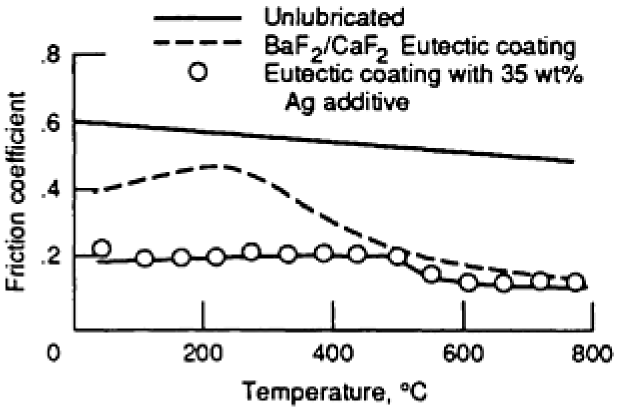


Fig. 22 Effect of temperature on friction coefficients of fused fluoride coatings (0.04 mm, or 0.0016 in., thick) sliding against cast Inconel pins with 4.76 mm (0.19 in.) hemispherical tip radius. Load, 5 N (1.1 lbf); speed, 2.3 m/s (7.5 ft/s). Source: Ref 49

Comparison of Ceramic-Bonded and Fused Fluoride Coatings. The ceramic-bonded coatings studied have good friction and wear properties at high surface temperatures that encourage the formation of a smooth, in some cases, vitreous, glaze on the sliding surfaces. They are not generally effective at low ambient temperatures except at high sliding velocities where frictional heating generates high surface temperature. Unfortunately, some vitreous glazes, which provide lubrication by viscous shear at high temperature, develop very high viscosity below their softening temperature. Below this temperature, the high viscosity of the glaze can result in very high bearing torque or even "lockup" of the bearing. A restart cannot occur until the bearing is heated above the softening point of the glaze. This phenomenon is not generally observed in pin-on-disk tests, where the contact area is very small and the adhesive bond is easily broken, but it has been observed in conforming-contact plane spherical bearings. This problem has not been observed with the nonvitreous glazes formed on the sliding surfaces of fused fluoride coatings that have no vitrifying oxides in their composition.

Plasma-Sprayed Coatings. When coatings are fusion bonded in a furnace, the entire part is heated above the melting point of at least one of the coating components. The high temperatures involved are not acceptable for some substrates, for example, titanium. An alternative method is plasma spraying, which only minimally heats the substrate. Plasma spraying is also more convenient (after the optimum plasma spray equipment parameters are determined), and can be used to deposit combinations of materials that are not amenable to furnace heat treatment. The first reported plasma spray tribological coatings developed at NASA Lewis were designated PS100 and PS101 (Ref 50). The former is a three-component composite coating of CaF_2, sodium-free, glass, and a nickel-chromium alloy binder. The PS101 designation also contains silver to improve low-temperature friction. Figure 23 shows the wear and friction-temperature characteristics of these coatings in low-speed, oscillating, plain spherical bearings. It is clear that the silver addition reduces friction at low temperatures without an adverse effect on lubrication at high temperatures. Both coatings are somewhat conformable, or ductile, and have moderately good wear characteristics. Coating and counterface wear factors are typically on the order of 10^{-5} $mm^3/N \cdot m$. PS100 and PS101 are used for high-temperature applications involving moderately loaded (up to at least 34 MPa, or 4900 psi) sliding contact bearings, and for shaft seals requiring a degree of conformability.

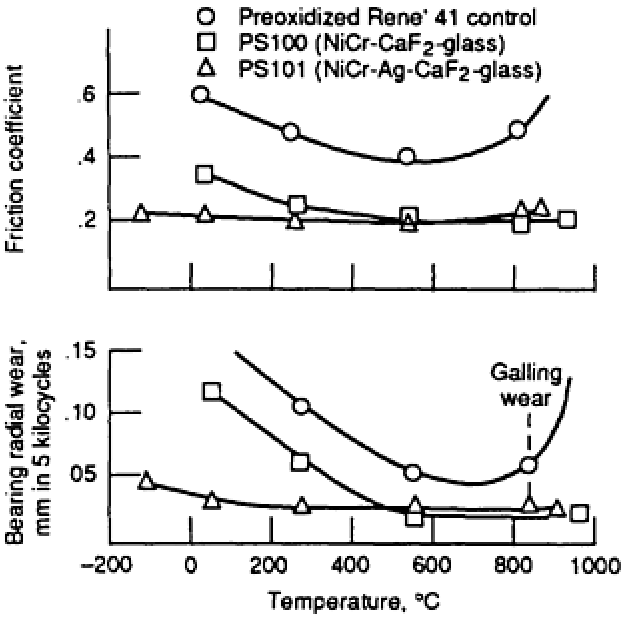


Fig. 23 Friction and wear properties of plasma spray coatings for self-lubricating oscillating (1 Hz, ±15°) bearings. Radial load, 34 MPa (5000 psi)

The PS200 series of plasma spray coatings were subsequently developed (Ref 51, 52). In this coating, the lubricating solids are distributed throughout a wear-resistant matrix of Ni-Co alloy-bonded chromium carbide (Cr_3C_2). The two most-studied compositions are PS200, which contains 10 wt% each of Ag and CaF_2-BaF_2 eutectic, and PS212, which contains 15 wt% each of the two lubricants in the bonded carbide matrix. The friction coefficients for these coatings are typically 0.30 to 0.35 from room temperature to 900 °C (1650 °F) at a sliding velocity of 2.7 m/s (8.9 ft/s) and tend to decrease at very high sliding velocities. Wear factors are about one-tenth of those for the PS100 series of coatings. In a composition-optimization study, PS212 had somewhat lower friction and wear in pin-on-disk tests than PS200; however, PS200 may be of interest where a harder coating is preferred.

Promising results have been obtained in a number of prototype tests of high-temperature mechanical components coated with PS200 or PS212. These include (1) start/stop tests of hydrodynamic gas bearings at temperatures up to 650 °C (1200 °F) (Ref 51, 53) and (2) a test of a four-cylinder Stirling engine with PS200-lined cylinders (Ref 51, 54).

Self-Lubricating Powder Metallurgy Composites. Plasma spray coatings are appropriate for flat surfaces and outside diameter surfaces that are readily accessible for spraying with a plasma spray gun. However, it is difficult or impossible to spray the inside surfaces of components such as small cylindrical bearings, valve guides, and bushings. These are the types of components that lend themselves to being made by powder metallurgy processes. The PS212 composition was used to make free-standing powder metallurgy (PM212) parts by sintering and by hot isostatic pressing (HIPing). The HIPed version is fully dense, whereas the sintered version is about 75% of theoretical density. The compressive strength at 900 °C (1650 °F) and the coefficient of thermal expansion of sintered PM212 and of HIPed PM212 are given in Fig. 24. The HIPed version is about three times stronger than the sintered version, but both materials retain compressive strength to at least 700 °C (1290 °F), thereby equalling or exceeding the room-temperature strength of many grades of bearing bronze and carbon sliding-contact bearing materials. The thermal expansion coefficient of PM212 is bracketed by those of such common alloys as cast iron and nickel-base superalloys in Fig. 24.

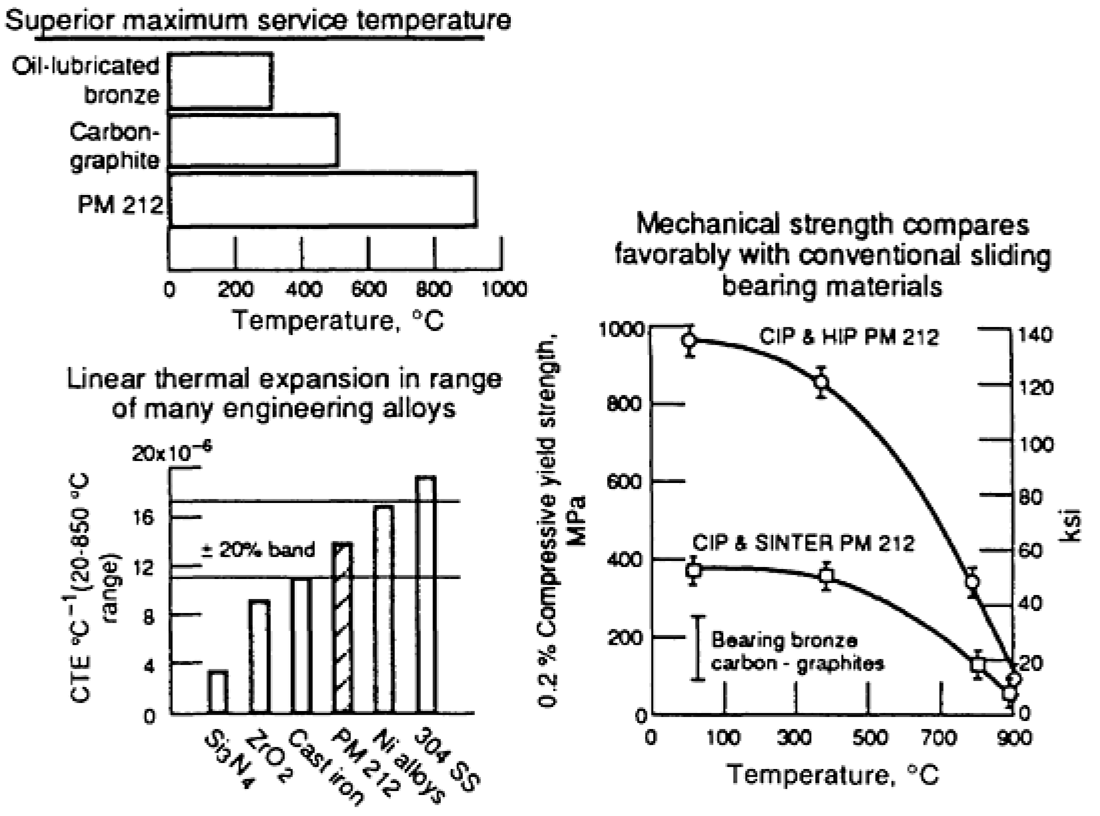


Fig. 24 Properties of PM212 self-lubricating P/M composite. Source: Ref 55

The tribological properties of sintered PM212 pins sliding on uncoated René 41 disks were reported by DellaCorte and Sliney (Ref 56). Figure 25 shows that friction coefficients at a sliding velocity of 2.7 m/s (8.9 ft/s) are about the same for sintered PM212 pins sliding on René 41 disks as they are for René 41 pins sliding on PS212-coated disks. The effect of sliding velocity on the friction coefficient of sintered PM212 is given in Fig. 26. At 760 and 900 °C (1400 and 1650 °F), the friction coefficient decreases with increasing sliding velocity and is 0.2 at the highest sliding velocity of 8.1 m/s (26.6 ft/s). Friction behavior is less straightforward at the lower temperatures, but friction coefficients are typically 0.30 to 0.35, except at the lowest velocity at room temperature. The trend of decreasing friction with increasing velocity suggests that friction coefficients considerably lower than 0.2 may be expected at the very high sliding velocities characteristic of shaft seals in turbomachinery. This effect may be due to the increased localized heating at high sliding velocities, which soften the silver and fluoride surface films, thus reducing their shear strength without appreciably reducing the hardness and flow pressure of the material below the near-surface region.

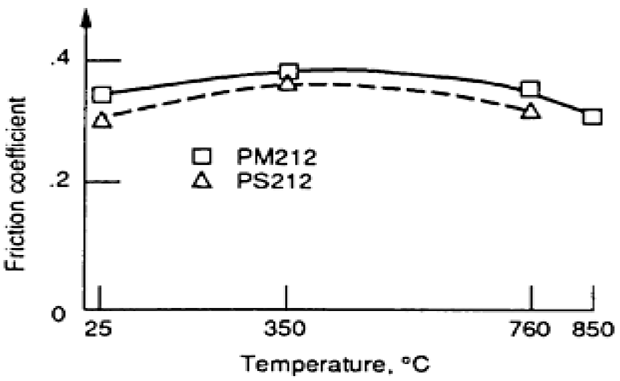


Fig. 25 Friction coefficient comparison of plasma spray coatings. Results from sintered PM212 pins sliding on René 41 disks and René 41 pins sliding on PS212 coated disks are given. Environment, air with 35% relative

humidity at 25% °C (75 °F); load, 0.5 kg (1.1 lbf); speed, 2.7 m/s (8.9 ft/s). Source: Ref 56

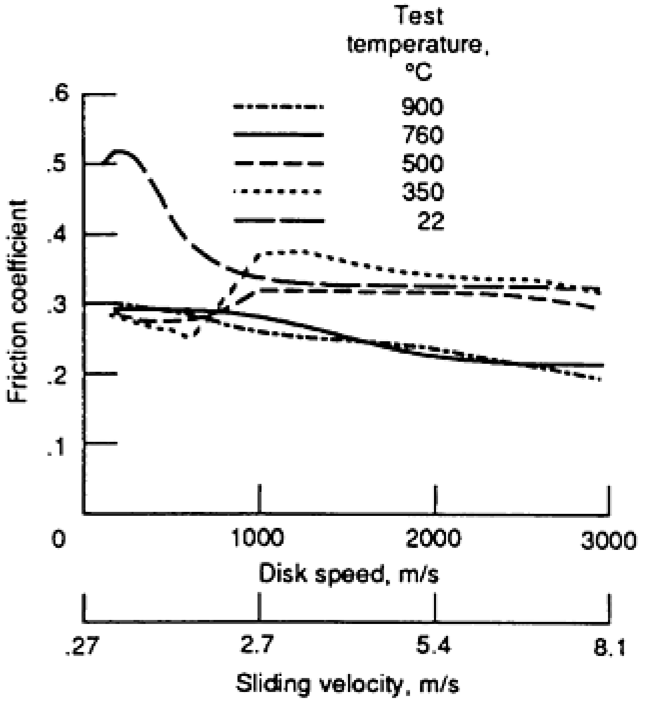


Fig. 26 Effect of sliding velocity on the friction coefficient of sintered PM212 sliding against René 41. source: Ref 56

Wear factors for sintered PM212 are in the 10^{-6} to 10^{-5} $mm^3/N \cdot m$ range. Wear factors for HIPed PM212 pins are about one-half those of sintered pins from 25 to 760 °C (75 to 1400 °F), and René 41 disk wear appears to be independent of whether the composite is sintered or HIPed.

References

1. R.R. Paxton, *Manufactured Carbon: A Self-Lubricating Material for Mechanical Devices,* CRC Press, 1979
2. S. Salto, *Elektrotech Z.,* 20 Sept 1906
3. W.O. Winer, Molybdenum Disulfide as a Lubricant: A Review of the Fundamental knowledge, *Wear,* Vol 10, 1967, p 422-452
4. H.S. Cooper and V.R. Damerell, "Lubricants Suitable for Various Uses," U.S. patent 2,156,806, 1939
5. M.E. Bell and J.H. Findlay, Molybdenite as a New Lubricant, *Phys. Rev.,* Vol 59 (No. 11), 1941, p 922
6. E.E. Bisson and W.J. Anderson, "Advanced Bearing Technology," NASA SP-38, National Aeronautics and Space Administration, 1964
7. H.E. Sliney, High Temperature Solid Lubricants, *Mech. Eng.,* Vol 96 (No. 2 and 3), 1974, p 18-22, 34-39; and H.E. Sliney, Solid Lubricant Materials for High Temperatures--A Review, *Tribol. Int.,* Vol 15 (No. 5), 1982, p 303-315
8. J.K. Lancaster, Solid Lubricants, *Handbook of Lubrication,* Vol II, *Theory and Design,* E.R. Booser, Ed., CRC Press, 1984, p 267-290
9. H.E. Sliney, Dynamics of Solid Lubrication as Observed by Optical Microscopy, *ASLE Trans.,* Vol 21 (No. 2), 1978, p 109-117
10. M.B. Peterson and R.L. Johnson, "Friction of Possible Solid Lubricants with Various Crystal Structures," NACA TN-3334, 1954
11. W.A. Brainard, "The Thermal Stability and Friction of the Disulfides, Diselenides, and Ditellurides of

Molybdenum and Tungsten in Vacuum," NASA TN D-5141, National Aeronautics and Space Administration, 1969

12. D. Godfrey and E.E. Bisson, "Bonding of Molybdenum Disulfide to Various Materials to Form a Solid Lubricating Film," NACA TN-2628, National Advisory Committee on Aeronautics, 1952
13. I. Ming Feng, Lubricating Properties of Molybdenum Disulfide, *Lubr. Eng.,* Vol 8, 1952, p 285-288, 306, 308
14. P. Fleischauer and R. Bauer, Chemical and Structural Effects on the Lubrication Properties of Sputtered MoS_2 Films, *Tribol. Trans.,* Vol 31 (No. 2), 1988, p 239-250
15. J. Lince, MoS_{2-x}-O_x Solid Solutions in Thin Films Produced by RF Sputter Deposition, *J. Mater. Res.,* Vol 5 (No. 1), 1990, p 218-222
16. D.H. Buckley, "Friction, Wear, and Lubrication in Vacuum," NASA SP-277, National Aeronautics and Space Administration, 1971
17. T. Spalvins and J.S. Przybyszewski, "Deposition of Sputtered Molybdenum Disulfide Films and Friction Characteristics of Such Films in Vacuum," NASA TM D-4269, National Aeronautics and Space Administration, 1967
18. T. Spalvins, Sputtering Technology in Solid Film Lubrication, *2nd International Conference on Solid Lubrication,* Society of Tribologists and Lubrication Engineers (ASLE), 1978, p 109-117
19. E.W. Roberts, Toward an Optimized Sputtered MoS_2 Lubricant Film, *20th Aerospace Mechanisms Symposium,* NASA CR 2423-Revised, National Aeronautics and Space Administration, 1986, p 103-119
20. H.E. Sliney, Decomposition Kinetics of Some Solid Lubricants Determined by Elevated Temperature X-Ray Diffraction Techniques, *Proceedings of the USAF Aerospace Fluids and Lubricants Conference,* P.M. Ku, Ed., Defense Documentation Center, 1963, p 350-367
21. R.H. Savage, Graphite Lubrication, *J. Appl. Phys.,* Vol 19 (No. 1), 1948, p 1-10
22. H. Kuwano and K. Nagai, Friction-Reducing Coatings by Dual Fast Atom Beam Techniques, *J. Vac. Sci. Technol. A,* Vol 4 (No. 6), 1986, p 2993-2996
23. A. Erdimir, G.R. Fenske, R.A. Erck, and C.C. Cheng, Ion-Assisted Deposition of Silver Films on Ceramics for Friction and Wear Control, *Lubr. Eng.,* Vol 46 (No. 1), 1990, p 23-30
24. T. Spalvins, "Bearing Endurance Tests in Vacuum for Sputtered MoS_2 Films," NASA TM X-3193, National Aeronautics and Space Administration, 1975
25. R.L. Fusaro and H.E. Sliney, Graphite Fluoride $(CF_x)_n$--A New Solid Lubricant, *ASLE Trans.,* Vol 13 (No. 1), 1970, p 56-65
26. A.K. Kuriakose and J.L. Margrave, Mass Spectrometric Studies of the Thermal Decomposition of Poly (Carbon Monofluoride), *Inorg. Chem.,* Vol 4 (No. 11), 1965, p 1639-1641
27. B.D. McConnell, C.E. Snyder, and J.R. Strang, Analytical Evaluation of Graphite Fluoride and Its Lubrication Performance Under Heavy Loads, *Lubr. Eng.,* Vol 33 (No. 4), 1977, p 184-190
28. R.L. Fusaro, Mechanisms of Graphite Fluoride $[(CF_x)_n]$ Lubrication, *Wear,* Vol 53, 1979, p 303-323
29. R.L. Fusaro, "Friction and Wear Life Properties of Polyimide Thin Films," NASA TM D-6914, National Aeronautics and Space Administration, 1972
30. R.L. Fusaro, Molecular Relaxations, Molecular Orientation, and the Friction Characteristics of Polyimide Films, *ASLE Trans.,* Vol 20 (No. 1), 1977, p 1-14
31. R.L. Fusaro and H.E. Sliney, Lubricating Characteristics of Polyimide Bonded Graphite Fluoride and Polyimide Thin Films, *ASLE Trans.,* Vol 16 (No. 3), 1973, p 189-196
32. R.C. Wagner and H.E. Sliney, Durable Solid Lubricant Coatings for Foil Gas Bearings to 315 °C, *3rd International Conference on Solid Lubrication,* Society of Tribologists and Lubrication Engineers (ASLE), 1984, p 105-112
33. R.L. Fusaro, Polyimides--Tribological Properties and Their Uses as Lubricants, *Polyimides,* Vol 2, K.L. Mittal, Ed., Plenum, 1984, p 1053-1080
34. H.E. Sliney and F.J. Williams, Performance of PTFE-Lined Composite Journal Bearings, *Lubr. Eng.,* Vol 39 (No. 8), 1983, p 511-518

35. J.P. Giltrow and J.K. Lancaster, Carbon-Fibre Reinforced Polymers as Self-Lubricating Materials, *Proc. Inst. Mech. Eng.,* 1968, p 149-159

36. J.P. Giltrow and J.K. Lancaster, The Role of the Counterface in the Friction and Wear of Carbon Fibre Reinforced Thermosetting Resins, *Wear,* Vol 16, 1970, p 359-374

37. R.L. Fusaro and H.E. Sliney, Friction and Wear Behavior of Graphite Fiber Reinforced Polyimide Composites, *ASLE Trans.,* Vol 21 (No. 4), 1978, p 337-343

38. H.E. Sliney and R.L. Johnson, "Graphite Fiber-Polyimide Composites for Spherical Bearings to 340 °C," NASA TN D-7078, National Aeronautics and Space Administration, 1972

39. H.E. Sliney and T.P. Jacobson, Performance of Graphite Fiber-Reinforced Polyimide Composites in Self-Aligning Plain Bearings to 315 °C, *Lubr. Eng.,* Vol 31 (No. 12), 1975, p 609-613

40. H.E. Sliney, Some Load Limits and Self-Lubricating Properties of Plain Spherical Bearings with Molded Graphite Fiber-Reinforced Polyimide Liners to 320 °C, *Lubr. Eng.,* Vol 35 (No. 9), 1979, p 497-502

41. H.E. Sliney, Evaluation of Two Polyimides and of an Improved Liner Retention Design for Self-Lubricating Bushings, *Lubr. Eng.,* Vol 41 (No. 10), 1985, p 592-598

42. M.N. Gardos and B.D. McConnell, "Development of a High-Load, High-Temperature, Self-Lubricating Composite," ASLE SP-9, Society of Tribologists and Lubrication Engineers (ASLE), 1982

43. R.L. Johnson and H.E. Sliney, Ceramic Surface Films for Lubrication at Temperatures to 2000 °C, *Am. Ceram. Soc. Bull.,* Vol 41 (No. 8), 1962, p 504-508

44. J. Lankford, W. Wei, and R. Kossowsky, Friction and Wear Behaviour of Ion Beam Modified Ceramics, *J. Mater. Sci.,* Vol 22 (No. 6), 1987, p 2069-2078

45. H.E. Sliney, T.N. Strom, and G.P. Allen, Fluoride Solid Lubricants for Extreme Temperatures and Corrosive Environments, *ASLE Trans.,* Vol 8 (No. 4), 1965, p 307-322

46. H.E. Sliney, "Rare Earth Fluorides and Oxides--An Exploratory Study of Their Use as Solid Lubricants at Temperatures to 1800 °F (1000 °C)," NASA TN D-5301, National Aeronautics and Space Administration, 1969

47. H.E. Sliney, "A Calcium Fluoride-Lithium Fluoride Solid Lubricant Coating for Cages of Ball Bearings to be Used in Liquid Fluorine," NASA TM X-2033, National Aeronautics and Space Administration, 1970

48. H.E. Sliney, An Investigation of Oxidation-Resistant Solid Lubricant Materials, *ASLE Trans.,* Vol 15 (No. 3), 1972, p 177-183

49. K.H. Olson and H.E. Sliney, "Additions to Fused-Fluoride Lubricant Coatings for Reduction of Low Temperature Friction," NASA TN D-3793, National Aeronautics and Space Administration, 1967

50. H.E. Sliney, Wide Temperature Spectrum Self-Lubricating Coatings Prepared by Plasma Spraying, *Thin Solid Films,* Vol 64, 1979, p 211-217

51. H.E. Sliney, A New Chromium Carbide-Based Tribological Coating for Use to 900 °C with Particular Reference to the Stirling Engine, *J. Vac. Sci. Technol. A.,* Vol 4 (No. 6), 1986, p 2629-2632

52. C. DellaCorte and H.E. Sliney, Composition Optimization of Self-Lubricating Chromium-Carbide-Based Composite Coatings for Use to 760 °C, *ASLE Trans.,* Vol 30 (No. 1), 1987, p 77-83

53. *Surf. Coat. Technol.,* Vol 36, 1988, p 87-97

54. D.J. Allen and W.A. Tomazic, "Hot Piston Ring Tests," NASA TM-100256, DOE/NASA 50112-72, National Aeronautics and Space Administration, 1987

55. P.E. Edwards, H.E. Sliney, C. DellaCorte, J.D. Whittenberger, and R.R. Martineau, "Mechanical Strength and Thermophysical Properties of PM212: A High-Temperature Self-Lubricating Powder Metallurgy Composite," NASA TM-103694, DOE/NASA/50162-5, National Aeronautics and Space Administration, 1990

56. C. DellaCorte and H.E. Sliney, "Tribological Properties of PM212: A High-Temperature Self-Lubricating, Powder Metallurgy Composite," NASA TM-102355, National Aeronautics and Space Administration, 1990

Grease

Raymond H. Boehringer, DuBois Chemicals, Inc.

Introduction

GREASE, in the form of mutton or beef fat (tallow), was used as early as 1400 BC to reduce axle friction in chariots. The semisolid consistency of fat allowed it to remain in place longer than liquids such as vegetable oil. Mutton and beef tallow were used until 1859, when Colonel Drake drilled his first oil well. Since that time, most lubricants have been based on petroleum (mineral) oil (Ref 1).

As defined by ASTM (Ref 2), lubricating grease is "a solid to semifluid product of dispersion of a thickening agent in a liquid lubricant. Other ingredients imparting special properties may be included." The function of a grease, according to Ref 3, is to remain in contact with moving surfaces and not leak out under gravity or centrifugal action, nor be squeezed out under pressure. Its major practical requirement is that it must retain its properties under shear at all temperatures to which it may be subjected during use. At the same time, the grease must be able to flow into the bearing through grease guns and from spot to spot in the lubricated machine as needed, but must not add significantly to the power required to operate the machine, particularly at the start. These parameters represent an exacting set of rheological requirements (Ref 3). Because of the solid nature of grease, it does not perform the cooling and cleaning functions of a fluid lubricant.

Greases are used in place of oils because (Ref 4):

- They stay put, either with less-expensive seals or no seals
- They reduce dripping and spattering
- They require less-frequent application
- They resist the washing action of water
- They seal out contaminants
- They can prolong the life of worn parts
- They provide a better mechanical lubrication cushion for extreme conditions, such as shock loading, reversing operations, low speeds, and high speeds

The lubricating grease industry has made a major contribution to the U.S. economy in the automotive, industrial, steel mill, farming, railroad, other transportation, and military sectors. Rugged industries, such as steel mills, test the maximum capability of new greases and the ingenuity of their manufacturers. Aircraft and military applications, which are beyond the capabilities of petroleum-based greases, utilize synthetic greases, which can sustain temperature extremes ranging from -54 to 315 °C (-65 to 600 °F).

Grease Basics

Grease consists of a thickening agent, base stocks, and additives (Fig. 1). Good sources for explanations of grease-related terminology are Ref 1 and 5. The base stock (oil) is the primary lubricant, although additives dramatically influence grease performance. Most thickeners do not contribute to lubrication, with the exception of calcium soap.

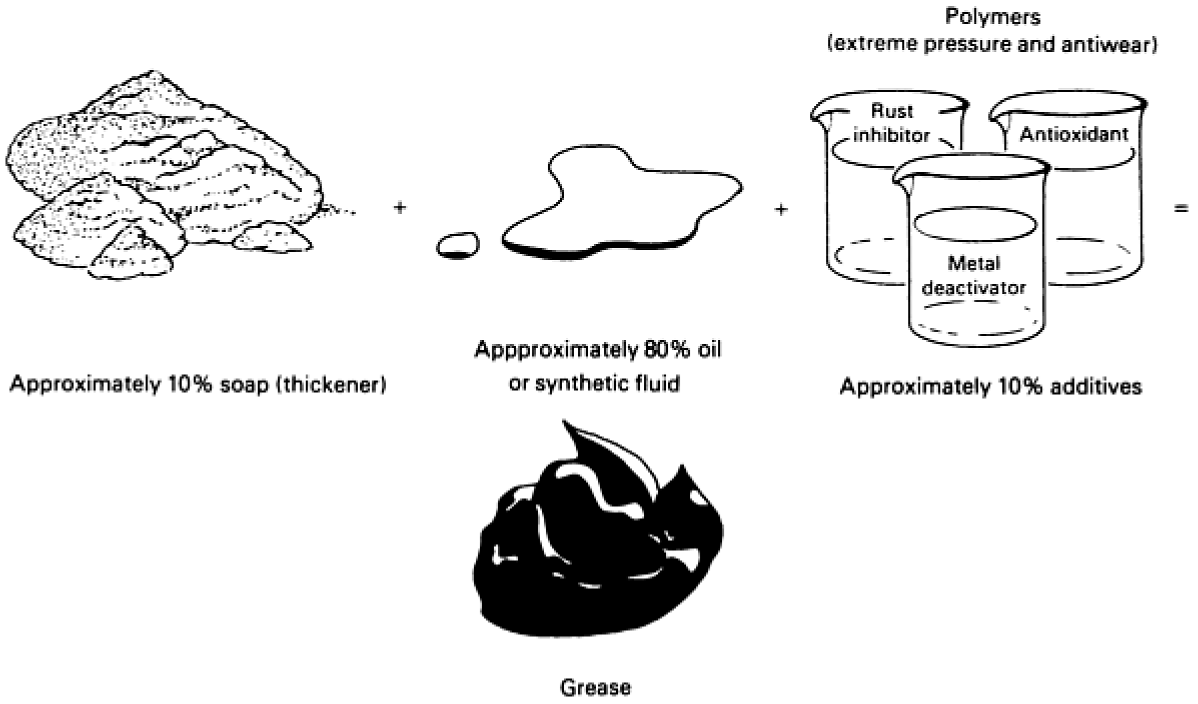


Fig. 1 Grease composition

Base Stocks

Base stocks can be either petroleum or synthetic liquid products; each category is described below.

Petroleum. Primary base stocks are petroleum based (mineral oil). Specifically, the stocks used in greases are petroleum oils separated from crude oil (Fig. 2), which is produced by oil wells. Crude oil is a black-colored mixture of fuels, petroleum lubricating oils of varying viscosities, and asphalt (tar-like products.) Fuels are a major part of crude oil and are maximized to 98 or 99% in quantity by chemical cracking.

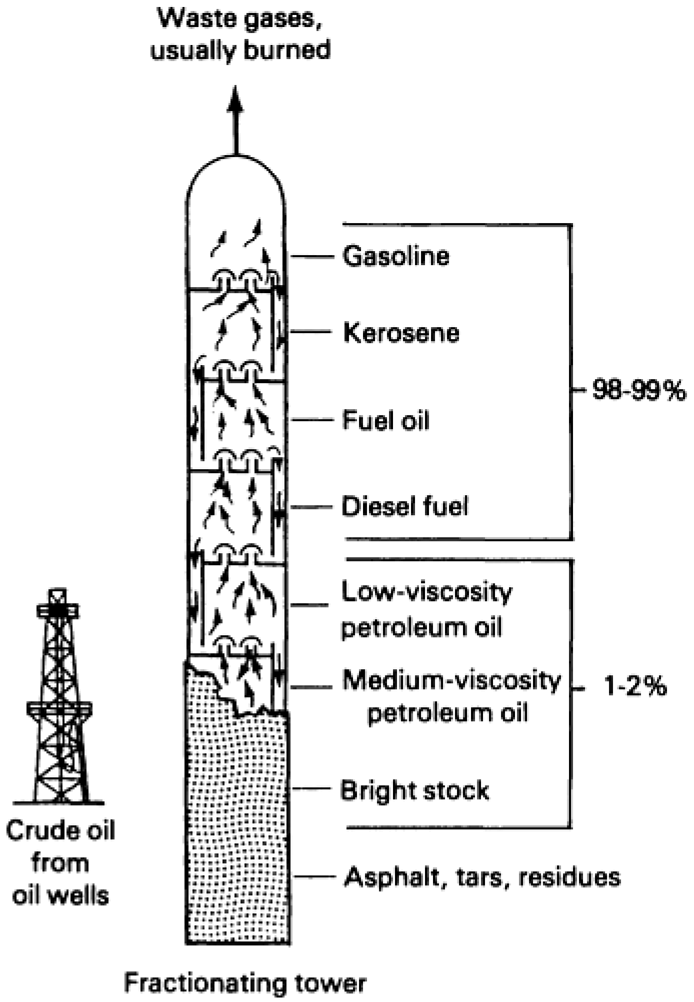


Fig. 2 Processing of crude oil by major oil companies

One of the major processes used to divide crude oil into its components occurs in the tall, thin fractionating towers associated with all petroleum refineries (Fig. 2). Flames are often seen at the top of these towers, because waste gases are frequently burned off. The petroleum oil fractions produced by this distillation tower can be used in grease manufacturing. The petroleum fractions range from light (low viscosity) petroleum to heavy, bright stocks (high viscosity). Low-viscosity fractions are used to impart better grease pumpability and good low-temperature performance. Most greases are made from medium-viscosity base stocks equivalent to an ISO 100 or ISO 150 (SAE 30 or 40) grade oil. Heavy-duty greases and high-temperature greases usually require heavier base stocks, approximately ISO 460 (SAE 140) grade.

Higher-quality petroleum-based stocks are used in higher-quality greases. Refining these base stocks can involve solvent extraction, bleaching, clay or carbon treating, hydrogenation, and redistillation. Water white-colored base stocks are referred to as white mineral oil, and are used in food grade, USDA-H1 greases that are suitable for incidental food contact.

Figure 3 shows the chemical structures of the three basic types of petroleum: paraffinic, naphthenic, and aromatic. Each type is described below, as well as in Table 1.

Table 1 Properties of three petroleum types that influence grease preparation and grease applications

Properties	Paraffinic	Naphthenic	Aromatic
Molecular form	Saturated	Saturated	Unsaturated
Oxidation stability	Best	Moderate	Poor
Thermal stability	Best	Moderate	Poor
Additive solubility	Poorest	Moderate	Best
Higher grease yields (use of less soap)	Poorest	Moderate	Best
Viscosity index	Highest	Low	Lowest
Low temperature	Fair	Good	Best
Flash points	Highest	Lower	Lowest
Boiling range	Highest	Lower	Lowest
Evaporation	Low	Moderate	Highest
Carbonization due to oxidation	Normally hard[a]	Soft	Soft
Toxicity	Minimal	Minimal	High
Elastomer seal swell[b]	Minimal	More swell	Most swell

(a) Can be corrected by better compounding.

(b) Oils do affect elastomers or rubber seals. Choice of seals with various greases must be considered. Selection of rubber-grease compatibility is beyond the scope of this article.

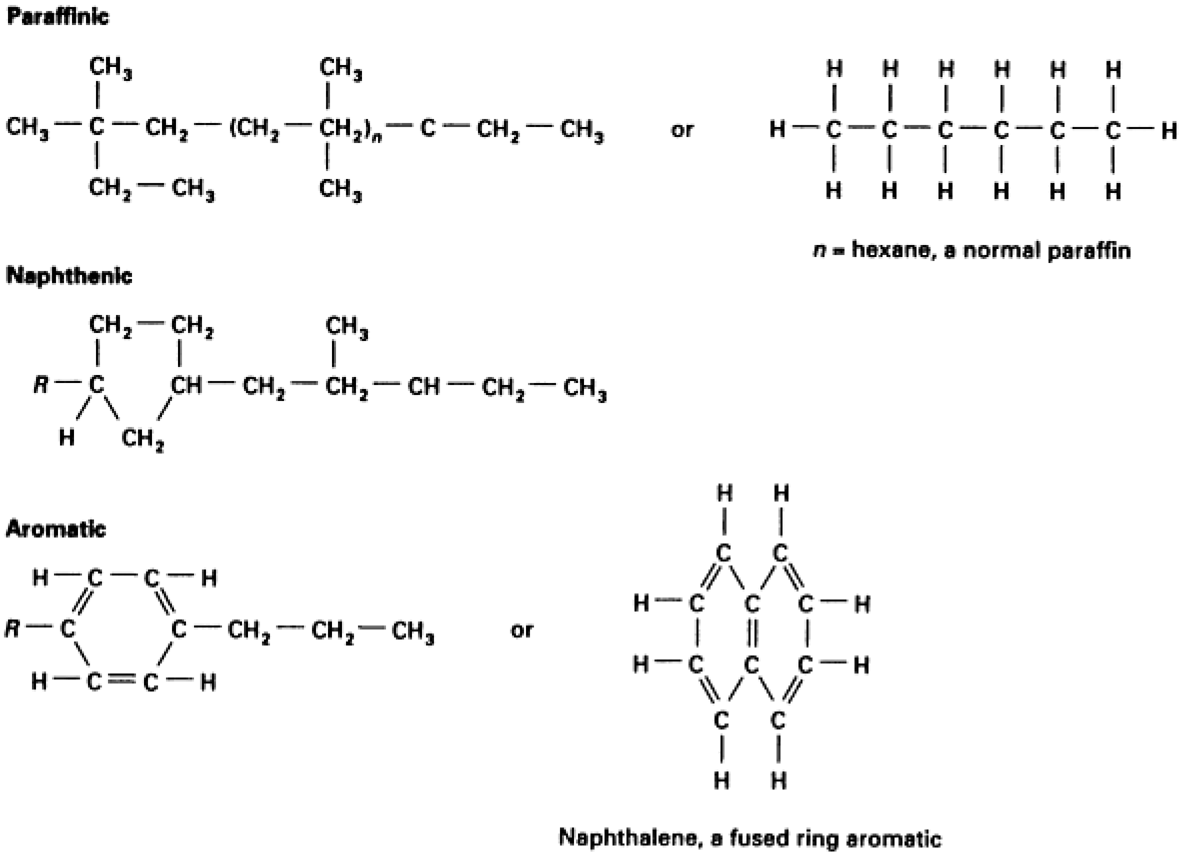


Fig. 3 Chemical structures of three basic types of petroleum: paraffinic, naphthenic, and aromatic

Paraffinic. The distinguishing feature of a paraffin is that the molecule is fully saturated; that is, no unsaturated, olefinic, double bonds exist. The carbons may be arranged in a straight line or branched.

Naphthenic stocks are similar to paraffinics, except that some carbon atoms are joined in saturated or cyclic ring structures.

Aromatic. Carbons in aromatic petroleum stocks are joined in hexagonal geometric structures with alternating double bonds between pairs of carbons. Often, two "aromatic" rings may be fused together. Use of aromatics are falling into disfavor, because of their toxicity. Aromatics can be thin oil (100-second), medium oil (30 weight), or heavy oil (bright stock, 140 weight).

Synthetics. A trend toward the use of synthetic lubricants began during the period from 1920 to 1945. The need arose for various reasons (Ref 6, 7). One reason was the increasing shortages of petroleum supplies in countries such as Germany, France, and Japan. Another reason was the demand for lubricants for new machines that required operability over greater temperature ranges than could be satisfied by petroleum. An early example was the liquid lubricant required for the jet engine. Both low- and high-temperature synthetic greases are used by industry and the military.

The major types of synthetic lubricant base stocks are:

- Esters
- Synthetic hydrocarbon (SHC) or polyalphaolefin (PAO)
- Polyglycols
- Silicones
- Chlorofluorocarbon
- Alkylated benzene
- Phosphate esters
- Polyphenyl ethers

Chemical structures are shown in Fig. 4, whereas Table 2 summarizes temperature ranges and typical uses of synthetic greases.

Table 2 Synthetic greases showing approximate temperature ranges and typical uses

Synthetic base fluid	Approximate temperature range[(a)]	Typical uses
Diesters	-68 to 176 °C (-90 to 350 °F)	Low temperature for aircraft and land-based vehicles (mostly military) and freezers. Also high temperatures
Polyolesters	-59 to 232 °C (-75 to 450 °F)	Low temperature for aircraft and land-based vehicles (mostly military) and freezers. Also, high temperatures
SHC synthetic hydrocarbon (PAO, polyalpha olefin)	-73 to 149 °C (-100 to 300 °F)	Industrial and automotive use, low temperature, high temperature to 150 °C (300 °F), freezers, can be food grade (USDA-H1)
Polyglycol	-34 to 176 °C (-30 to 350 °F)	Lubrication of rubber or plastic on steel. Resistance to chlorine
Silicones	-73 to 260 °C (-100 to 500 °F)	Lubrication of rubber or plastic on steel. Good temperature extremes. Poor lubricant for steel on steel. High-vacuum grease
Chlorofluorocarbon	-51 to 315 °C (-60 to 600 °F)	Chemically inert to oxygen, chlorine, acids, nonflammable. Good for rubber and plastics
Alkylated benzene	-26 to 176 °C (-15 to 350 °F)	Low-temperature applications. Radiation resistant
Phosphate esters	-26 to 260 °C (-15 to 500 °F)	Fire resistance
Polyphenyl ethers	10 to 315 °C (50 to 600 °F)	Very high vacuum diffusion pumps. High temperatures. Radiation resistant

(a) Temperature range of base fluid, which can be maintained in grease form by proper selection of thickener

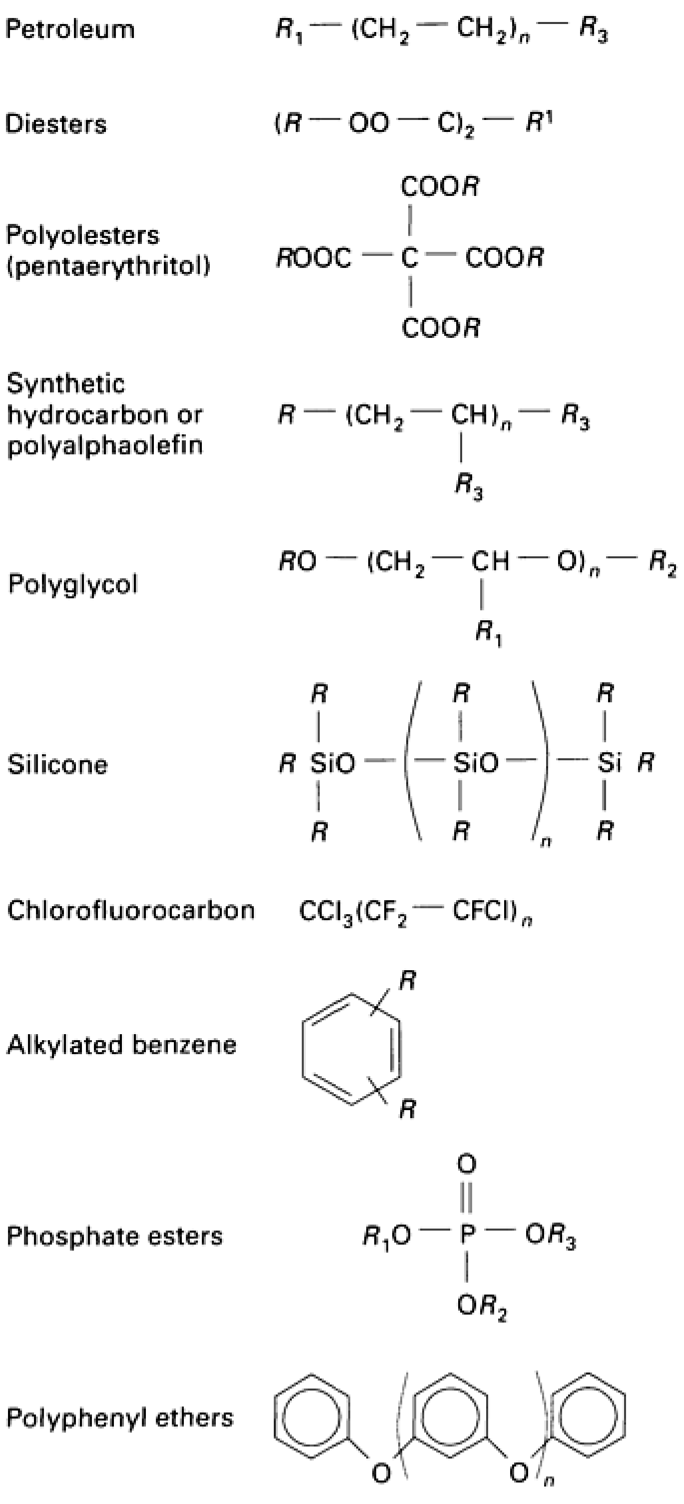


Fig. 4 Chemical structures of synthetic lubricant base stocks versus petroleum

Synthetic lubricants are manmade products that are used when petroleum-based oils can no longer perform. Conditions for use include high temperatures (95 to 315 °C, or 200 to 600 °F) and low temperatures (-40 to -75 °C, or -40 to -100 °F), and when there is a need for longer lubricant life or clean operation with less carbon, sludge, and varnish.

Additives (Ref 8)

The basic additives for making grease are:

- Antioxidants, for longer grease life because of better resistance to oxidation from air at high temperatures
- Anticorrosion agents, to provide resistance to chemical attack on alloy bearings
- Rust inhibitors, to protect against rust formation under wet conditions
- Extreme-pressure agents, for high film strength under shock loads, such as gears reversing direction rapidly, worm gears, or hypoid gears
- Solid additives, such as molybdenum disulfide (moly) and graphite, are *sometimes* used for back-up lubrication in high-temperature applications over 315 °C (600 °F). Solid additives also have some application in greases that need extreme high-pressure properties, such as roll neck bearings in a steel mill. Without adherence to proper greasing cycles, these solids can accumulate in bearings, causing wear. Either heat or too-infrequent greasing can cause oil evaporation or bleed, concentrating thickener and solids
- Polymers or viscosity index improves, which are used to prevent both leakage of grease out of bearings and the entrance of contaminants into the bearing. Polymers improve lubrication properties, control the rate of oil bleed to allow the grease to last longer, and improve water resistance

Types of Grease-Thickener Systems

Lubricating grease is a solid to semifluid mixture of a fluid lubricating oil and a thickening agent. A typical grease consists of 90% oil and 10% thickening agent. Table 3 (Ref 8) shows thickener types, normal percent thickener, melting (dropping) point, and typical uses. The National Lubricating Grease Institute (NLGI) grade is defined in specification IID, grease classifications.

Table 3 Effect of the thickener type on melting point and typical grease use

Total U.S. grease market, %	Thickener type	Thickener for NLGI grade 2, %	Dropping point[(a)]	Typical uses based on properties of thickener[(b)]
50.12	Lithium soap	7.25	188 °C (370 °F)	Most widely used grease; can be pumped in central grease systems
15.28	Lithium complex soap	12	>260 °C (>500 °F)	Multipurpose; high-temperature; central systems
2.97	Sodium soap	12	182 °C (360 °F)	Water soluble grease; low-cost "old" wheel bearing grease; fibrous
4.58	Calcium soap, hydrated	14	88 °C (190 °F)	Low-cost cup and water pump grease
3.99	Calcium soap, anhydrous	7.25	143 °C (290 °F)	Multipurpose; central systems; good water resistance
3.02	Calcium complex soap	16	>260 °C (>500 °F)	Multipurpose; good water resistance; good high temperature; NLGI 0 grade needed to pump in central systems
New	Calcium sulfonate	20	>260 °C (>500 °F)	Heavy-duty high extreme pressure; high temperature
6.07	Aluminum complex soap	6	249 °C (480 °F)	Best water resistance; best chemical resistance
0.88	Other soaps	9-30	Variable	. . .
1.3	Other thickeners, nonsoap, organic, inorganic	9	. . .	. . .
5.03	Polyurea	9	>260 °C (>500 °F)	Multipurpose; central systems; wide temperature range; long life; good low temperature
6.34	Benton (clay)	7	>260 °C (>500 °F)	Multipurpose; excellent high temperature (does not melt)

Source: Ref 8, p 26

(a) Dropping point of grease is that temperature where it melts and becomes liquid.

(b) Consider that base stock selection also has some effect on use. See Table 2 for synthetics. Refer to text on petroleum base stocks.

The manufacture of a grease is still an art. There are a limited number of grease manufacturing plants in the United States and throughout the world. Scientific understanding of grease manufacturing is increasing, but has not replaced the art that is necessary.

The thickening agent can be thought of as a sponge. This sponge is usually soap, which is crystalline or micellular in nature. The oil portion saturates the sponge. Typical chemical structures of soaps are shown in Fig. 5. The simple soaps have a lower dropping point. Complex soaps are larger molecules that allow the grease to have a higher dropping point.

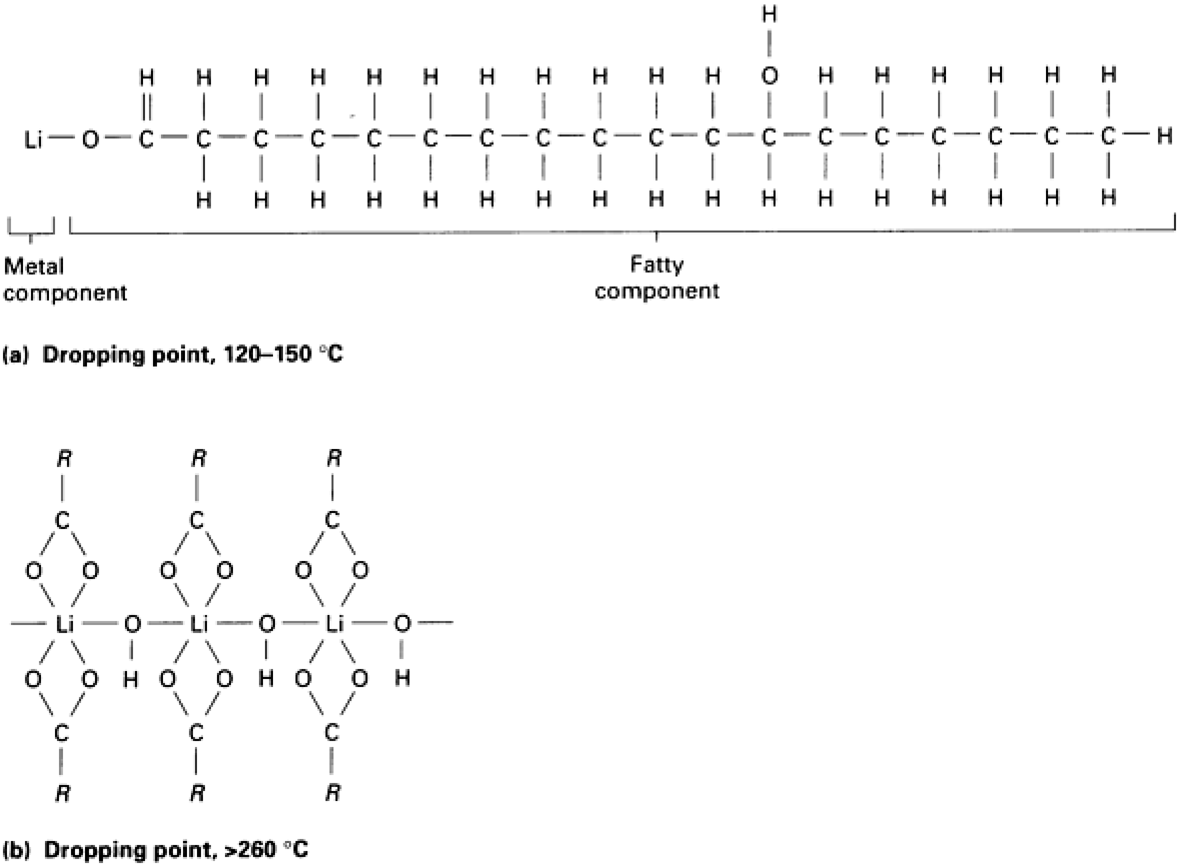


Fig. 5 Typical chemical structures of soap thickeners. (a) Simple soap. (b) Complex soap, proposed structure (lithium 12-hydroxystearate). Source: Ref 3

Table 4 summarizes thickener types and provides reasons for their use. Most thickeners are soaps that have a metal component, which affects grease performance. The fatty component of each soap type is usually 12-hydroxystearic acid.

Table 4 Thickener types and reasons for use

Thickener	Reasons for use
Soaps (metal component)	
Lithium	Most common grease in the world; easy to manufacture; easy to store; good pumpability; resistance to dust and coal; its flowability allows contaminants to flow out
Calcium	Requires less regreasing; good water resistance; the soap itself aids lubrication
Aluminum	Highest resistance to water, chemicals, acids, bases
Barium	High water resistance; somewhat toxic
Sodium	Fibrous; water soluble
Inorganics, clays, silica	No dropping point; does not melt
Polyurea (organic)	High water resistance; pumpable, long life, good at low temperature

Grease Classifications

The NLGI has classified the consistency of greases on the basis of ASTM D 217, Cone Penetration Test (refer to NLGI specification IIE, grease testing). Penetration is the consistency, or hardness, of a grease measured by the distance that a standard cone will penetrate the surface after a free fall at a standard temperature. Table 5 identifies NLGI grades and penetration measurements.

Table 5 NLGI grease classifications and consistency grades

NLGI grade	ASTM penetration at 25 °C (77 °F), 10^{-1}mm	Consistency
000	445-475	Semifluid
00	400-430	Semifluid
0	355-385	Semifluid
1	310-340	Soft
2	265-295	Common grease
3	200-250	Semihard
4	175-205	Semihard
5	130-160	Semihard
6	85-115	Hard

Common grease is NLGI grade 2. It is normally sold in cartridges for ordinary grease guns.

Thin greases, which are less than NLGI grade 2, pump more easily in central lubricating systems and generally allow operation at lower temperatures. They are sometimes used in cases of high (above 5000) rev/min.

Heavier greases, which are greater than NLGI grade 2, have limited use. A classic use of grade 6 block grease is in a feed box for journal bearings in old paper mills. Block, or sock, greases were once used for railroad journal wheel bearings. The new American Association of Railroads specification 942-88 has switched to large, tapered roller bearings for railroad cars, for which NLGI grade $1\frac{1}{2}$ nonextreme-pressure grease is used. Locomotives use the extreme-pressure version.

Heavy greases are sometimes used for high-speed bearings above 5000 rev/min. The high-speed theory is that the heavy grease channels allow the bearing to rotate at a high rev/min rate, but a trace amount of grease contacts the bearing, keeping it lubricated.

Grease Testing

The most frequently used grease laboratory testing procedures are those of ASTM (Ref 9, 10), which are summarized in Table 6.

Table 6 Summary of grease laboratory tests

Test name	ASTM No.	Test description	Results
Penetration	D 217	Consistency or hardness of a grease	Range from semifluid to brick-hard grease
Shear stability	D 217A	Ability of grease to withstand repeated working with minimum change in consistency	Results after 10,000 strokes: 5% excellent; 5-15%, good; 15-30%, fair; >30%, poor
Dropping point	D 566	Temperature at which grease melts or turns liquid	Typical drop points: simple soaps, 120-175 °C (250-350 °F); complex soaps, 205-290 °C (400-550 °F); nonsoap, no melting point
Shell roller test	D 1831	Simulates conditions of roller bearing greases	Penetration change after 4 h of rolling: 0-10% excellent; 10-25%, good; 25-60%, fair; >60%, poor
Bleeding and evaporation	U.S. Steel	Measures oil bleed of greases in storage	Held at 100 °C (212 °F) for 50 h
Wheel bearing test	D 1263	Measures leakage and varnish deposits	Held at 113 °C (235 °F) for 6 h; target is zero

			loss and no varnish
Extreme pressure (Timken)	D 2509	To determine maximum load the bearing or gear will carry	Greater than 155-180 N (35-40 lbf) "OK load" is usually considered an extreme-pressure grease
Cloud and pour points of base stock	D 97	"Cloud" temperature at which solids begin to crystallize; "pour," temperature when there is no movement for 5 s	Low-temperature performance

Applications

This section describes the parameters that should be considered when selecting a grease, and discusses the incompatibility of greases.

Grease Selection

Although the applications described in this article are industrial in nature, the same rules apply when selecting grease for automotive service. Interested readers are referred to ASTM D 4950, "Classification of and Specifications for Automotive Service Greases." Consideration of the parameters described below is necessary for engineers or maintenance personnel who must select the best grease for a particular job.

Bearing Temperature. The grease dropping point should be about 27 °C (50 °F) above the bearing temperature so that the grease does not liquify. The evaporation rate of the base stock at various high temperatures can be used as a parameter for choosing a grease. If the evaporation rate is high, then the grease may be unsuitable or may require many applications.

A laboratory test for evaporation is ASTM D 2595, "Evaporation Loss of Grease," which measures evaporation loss of lubricating greases at temperatures between 95 and 315 °C (200 and 600 °F). Evaporation should be low (5 to 10%, maximum) at the use temperature. The higher the evaporation loss, the more frequently grease will need to be replaced. Evaporation that is too fast will cause grease hardening because of lack of oil and increased concentration of thickener.

Low Temperature. As an indicator for low-temperature applications, the pour point of the grease base stock is used as the minimum temperature at which the grease will not be frozen. Thinner greases, such as NLGI grades 0 or 00, pump more easily at low temperatures than does a normal NLGI grade 2 grease.

Several low-temperature grease tests are available to help define low-temperature grease behavior:

- U.S. Steel, grease mobility testing, usually run at -18 °C (0 °F)
- ASTM D 4693 (1987), "Low Temperature Torque of Grease Lubricated Wheel Bearings"
- ASTM D 1478, "Low Temperature Torque of Ball Bearing Greases," which was developed specifically for greases with low torque at -54 °C (-65 °F) for aerospace activities
- ASTM D 1092, "Apparent Viscosity of Lubricating Greases," in which viscosity is measured from -54 to 38 °C (-65 to 100 °F)

Speed. Using conventional greases, the speed (rev/min) of a bearing or gear is usually not a problem. Most greases can easily handle 3500 rev/min. The normal speed of electric motors is 1750 rev/min. Normal industrial equipment runs at less than 1750 rev/min. Some large grease manufacturing companies (Ref 11) use ASTM D 3336 to test their greases on lab equipment that operates as high as 10,000 rev/min. At these higher speeds, most greases still perform, but the equipment can get noisy.

The choice of grease can also be aided by using a DN value. This value, also called the speed factor, is the product obtained by multiplying the diameter of a bearing by the number of rev/min, that is, DN = diameter (in mm) × N (rev/min). The shaft diameter should be used for journal (sleeve) bearings, and the pitch diameter, for antifriction bearings. For a diameter given in inches, multiplying by 25.4 will yield mm. Table 7 identifies grease selection bases on DN values.

Table 7 Selection of NLGI grade grease at various *DN* values

For *DN* values of:	NLGI grade No.
0-50,000	3
50,000-125,000	2
125,000-250,000	1
250,000-350,000	0

Note: This general *DN* guide holds for multipurpose greases whose oil viscosity is below 250 mm^2/s at 40 °C (1200 Saybolt universal seconds at 100 °F.) Lower *DN* values may have to be used with greases having either tacky properties or higher oil viscosities.

Pressure. Although most industrial equipment does not require extreme-pressure greases, it is advisable to use greases that contain these additives. They will protect the bearings or gears when either extreme-pressure or shock-loading conditions arise.

Standard laboratory tests that are used to measure extreme-pressure greases are:

- ASTM D 2509, "Standard Method for Measurement of Extreme-Pressure Properties of Lubricating Grease (Timken Method)," where the minimum load for extreme-pressure greases should be 178 to 223 N (40 to 50 lbf)
- ASTM D 2596, "Standard Method for Extreme-Pressure Properties of Lubricating Grease (Four-Ball Method)," where the minimum weld point for extreme-pressure greases should be about 1.96 kN (200 kgf)

The weld point, under the conditions of the test described above, is the lowest applied load in kgf at which the one rotating ball seizes and then welds to the three stationary balls, which indicates that the extreme-pressure level of the grease has been exceeded.

Viscosity, Speed, and Pressure (Ref 12). The relationship of these parameters is of interest as a point of reference. The base oil viscosity of a grease is the primary determining factor in its ability to provide a proper lubrication film. The relationship that exists for the oil portion of the grease is: coefficient of friction = ZN/P, where Z is the viscosity of the oil, N is the speed in rev/min, and P is the pressure. To understand this formula, consider that most high-speed operations are associated with low-viscosity oil and light loads, and that most low-speed operations are associated with high-viscosity oil and heavy loads.

Viscosity as Function of Bearing Diameter and Speed. A method for choosing the minimum viscosity base oil required in a grease to adequately lubricate the bearing at the operating temperature has been devised (Fig. 6 and 7).

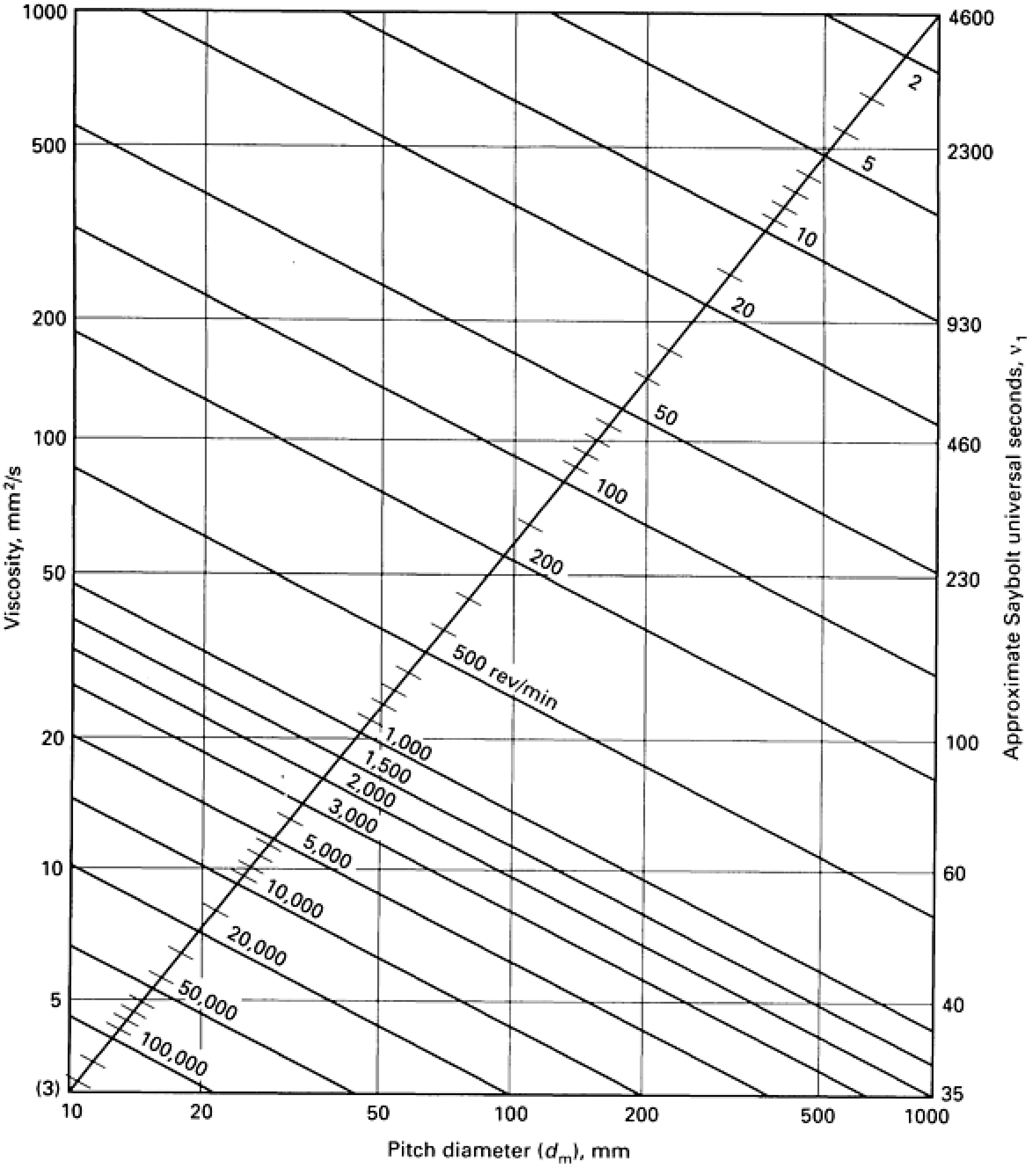


Fig. 6 Minimum required lubricant viscosity as function of pitch diameter and speed. d_m = (bearing bore + OD) ÷ 2; v_1 = required lubricant viscosity for adequate lubrication at operating temperature

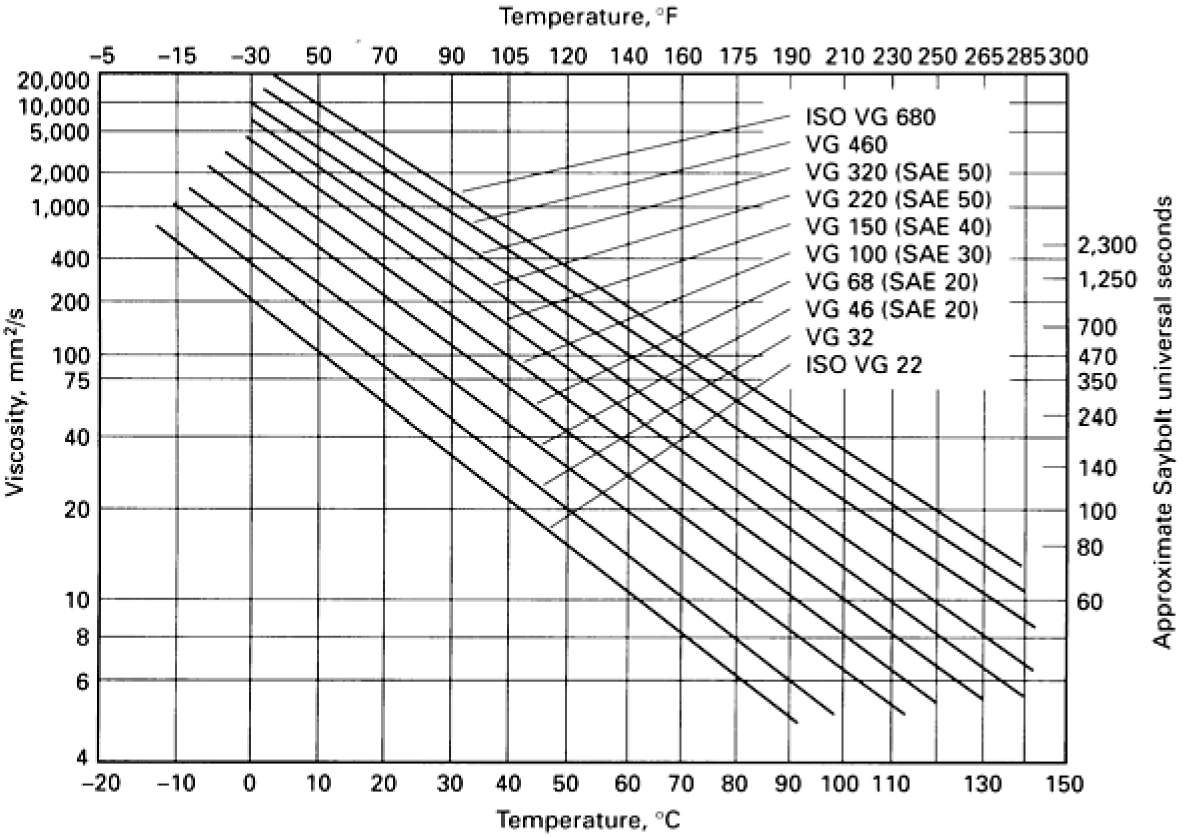


Fig. 7 Viscosity-temperature chart. Viscosity classification numbers per ISO 3448-1975 for oils having viscosity index of 95; approximate equivalent SAE viscosity grades shown in parentheses

Conditions, Environment, and Contamination. Most industrial greases are not subjected to unusual environmental conditions, because they are used inside plants with clean, controlled conditions. Grease selection can thus be made with more freedom, based on other parameters. However, Table 8 identifies unusual or extreme environmental conditions and chemicals, and recommends the use of specific greases. The requirements associated with hot and cold temperatures were discussed in the sections "Bearing Temperature" and "Low Temperature" in this article.

Table 8 Grease selection under adverse conditions

Unusual condition	Suggested grease
Water, rolling oil, cutting fluid	Aluminum-complex grease should be used where large quantities of water run over the bearing or gear; calcium-complex grease is effective for more incidental water contact
Coal dust, paper particles, dust, foundry atmosphere	Lithium-complex grease or grease with good flow characteristics and low-viscosity base stock to allow dust to work out; NLGI grades 1 or 2
Presence of acids, bases	Calcium sulfonate [over-based[a]] is particularly good in the presence of acids; petrolatum (vaseline); aluminum complex
Harsh chemicals, such as chlorine and oxygen	Grease made from chlorofluorocarbon base stock. *Caution: Do not use grease made from petroleum, because it will burn*
Rubber on steel lubrication	Greases from polyglycol or silicone base stock

(a) Overbased sulfonate detergents were used in the early 1950s in automotive engine oils. It was found that fully oil soluble sulfonates could be prepared using 3 to 15 times as much metal, calcium in this case, as the normal sulfonates. Highly based sulfonates were found to possess a great capacity to neutralize acidic contaminants.

Pumpability is an important property because grease must reach the bearing or gear that needs lubrication. Because grease guns develop extremely high pressures, they can pump almost any grease directly into a bearing. Small-diameter lines between the grease gun and a bearing can be a problem. The grease must be carefully selected when small lines and

centralized grease systems are used. Three tests (Table 9) can aid in this selection. Factors involved in the pumpability of grease are described below.

Table 9 Tests to evaluate grease pumpability in central systems and small grease lines

Test	Description
Apparent viscosity of lubricating grease, ASTM D 1092	Greases are nonNewtonian in nature and may be said to have an "apparent viscosity" (not the true viscosity). The apparent viscosity holds only for the shear rate (and temperature) at which the viscosity is determined. The test involves forcing the grease through a capillary with a piston
Lincoln ventmeter	Apparatus consists of a 7.5 m (25 ft) coil of $\frac{1}{4}$ in. copper tubing. Grease is pumped into the coil and pressurized at 12.4 MPa (1.8 ksi). The grease is released and the pressure measured at the end of 30 s. The test can predict grease line diameter and length
Trabon or Lubriquip Houdaille grease test	Lubriquip runs a grease in its commercial central lubricating systems and selects the appropriate system to pump grease

Fluid Viscosity. Greases made with high-viscosity base stocks are harder to pump than those made with thin base stocks.

Temperature. Warm greases pump easier. Low temperature has been discussed previously.

Thickener type was described in Tables 3 and 4. The classical central lubricating greases are lithium thickened. Although greases that use other thickeners can be pumped in central systems, grease composition may have to be adjusted by using either thinner base stocks or less thickener.

Amount of thickener. Grease with an NLGI grade 0 rating can be pumped easier than an NLGI grade 2 grease of the same type. In any family of greases, the NLGI grade is dependent on the amount of thickener.

A centralized grease system, described below in "Dispensing Methods," is the best method for supplying grease to a large number of bearings on a machine. A bearing constantly uses grease in small quantities, and a central system can supply small quantities as often as desired. A centralized grease system is also designed for operation within a definite pressure range. Exceeding this pressure can cause the separation of thickener and base stocks, which then plug the system. Therefore, grease pumpability is crucial.

Seals are important to the grease system because they keep the grease in and contaminants out. Parameters that need to be considered are described below.

Composition. Normal seals used in industry include Buna-N, neoprene, viton, and silicone rubbers.

Compatibility of the seal with the grease being used is important. The rubber can be tested for changes in volume, appearance, cracking, Durometer hardness, elongation, and tensile strength. Test methods include ASTM D 4289, "Compatibility of Lubricating Grease with Elastomers," established in 1983, and the older federal test methods (FTM) 791e, 3603.4, and 3604.1, run at temperatures from 70 to 205 °C (158 to 400 °F).

Temperature. It is important to determine the maximum and minimum temperatures that the seal will encounter, and to select the seal accordingly.

Paints. If paints are encountered either inside or outside the grease system, they should be tested for compatibility with the grease. Because there are so many paints and surface coatings, it would be most expedient to contact the paint company for general guidelines and/or tests.

Incompatibility of Greases

The compatibility of greases is largely controlled by the thickener type (Table 10). If the thickeners of two greases are the same, then they will probably be compatible. Greases made from different thickeners are generally incompatible, because of the chemical reaction between the thickeners. Compatibility problems arise primarily when grease is supplied through a gun into a zerk fitting, and then into the bearing. To avoid problems, the grease gun should be slowly pumped until the grease being exhausted from the bearing is the same color as the grease being pumped. This purging procedure normally

replaces about 90% of the old grease. If only about 50% of the grease is replaced, then the grease mix usually liquifies and runs out of the bearing in 5 or 10 min. Grease compatibility problems do not seem to arise if dissimilar greases are fed to bearings through central systems. The exact reason is not known.

Table 10 Grease incompatibilities

	Aluminum complex	Barium	Calcium	Calcium 12-hydroxystearate	Calcium complex	Clay	Lithium	Lithium 12-hydroxystearate	Lithium complex	Polyurea
Aluminum complex	. . .	I	I	C	I	I	I	I	C	I
Barium	I	. . .	I	C	I	I	I	I	I	I
Calcium	I	I	. . .	C	I	C	C	B	C	I
Calcium 12-hydroxystearate	C	C	C	. . .	B	C	C	C	C	I
Calcium complex	I	I	I	B	. . .	I	I	I	C	C
Clay	I	I	C	C	I	. . .	I	I	I	I
Lithium	I	I	C	C	I	I	. . .	C	C	I
Lithium 12-hydroxystearate	I	I	B	C	I	I	C	. . .	C	I
Lithium complex	C	I	C	C	C	I	C	C	. . .	I
Polyurea	I	I	I	I	C	I	I	I	I	. . .

B, borderline compatibility; C, compatible; I, incompatible.

Source: Ref 13

Practices

The practices described in this section include the methods and frequency of grease application, as well as the problems associated with either overgreasing or undergreasing. The storage and handling of greases are also discussed.

Dispensing Methods (Ref 10)

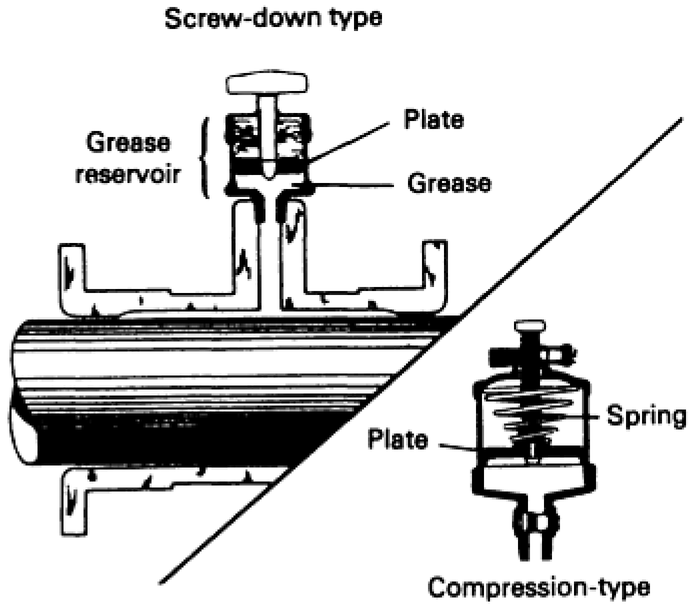
There are basically three methods of applying grease: by hand; by hand- or air-operated mechanical devices, which deliver grease to one point of use at a time; and by centralized grease systems, which supply a number of points of use from one central reservoir.

Hand application involves the direct application of grease to the part. This method is frequently used during the assembly of a machine. Grease is spread over bearings and gears to protect them from rust and to ensure lubrication when the machine is started for the first time. Balls and roller bearings require that grease be worked by the fingers into the spaces between the balls or rollers, in addition to requiring an overall coating. However, dirt is a predominating threat in hand greasing.

Another hand method involves pouring melted grease from a suitable container onto gears or open guide bearings. In any hand method, care should be taken not to apply too much grease, because the excess is squeezed out and wasted.

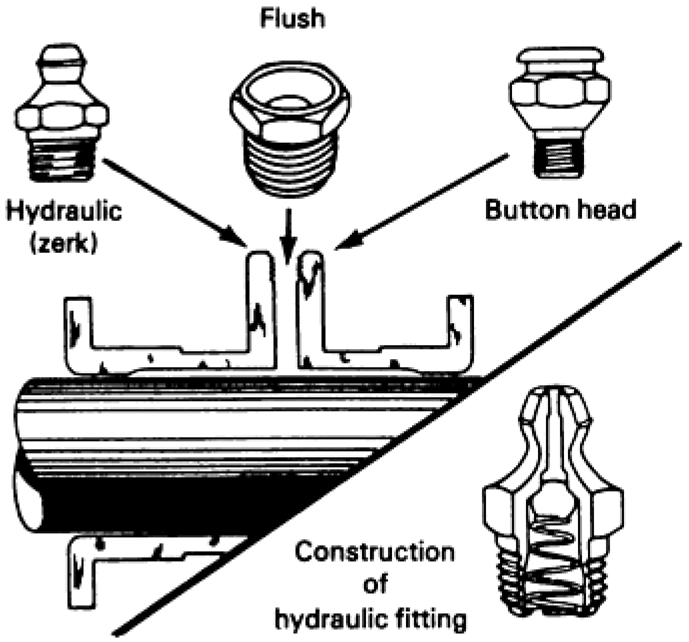
Hand- or air-operated mechanical devices deliver grease to one point of use at a time. A variety of devices are described below.

Grease cups are used to apply grease to bearings. The ordinary screw-down type consists of a small reservoir for holding the grease and a plate that screws down into the reservoir, exerting pressure on the grease and forcing it into the bearing. It is filled by unscrewing the plate, filling the reservoir with grease, and replacing the plate.



The compression type of cup has a spring-loaded plate, which maintains pressure on the grease and forces it slowly into the bearing. Compression types with a spring are now available with a zerk grease fitting for refilling.

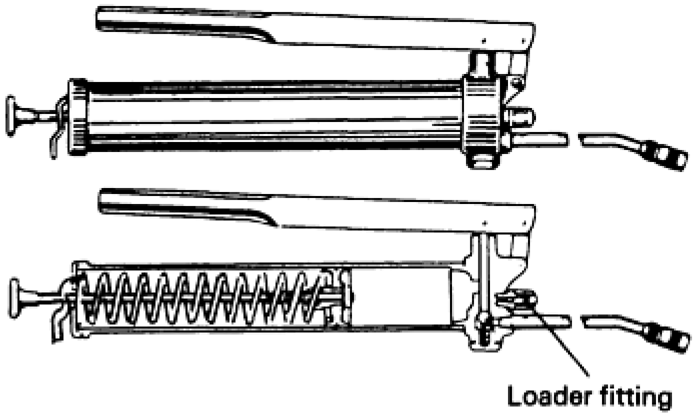
Grease fittings, which make possible the use of pressure grease guns, have replaced grease cups on most steel-mill equipment, because they reduce the time needed for greasing. Of the several types available, the hydraulic fitting is most frequently used on smaller and lighter machinery. A "flush" fitting is used when clearance or other design considerations warrant it. Although it is not actually flush, it extends very little above the surface of the bearing housing. Button-head fittings are used on heavier mill equipment, where severe conditions prevail and large volumes of grease must be forced through the fitting.



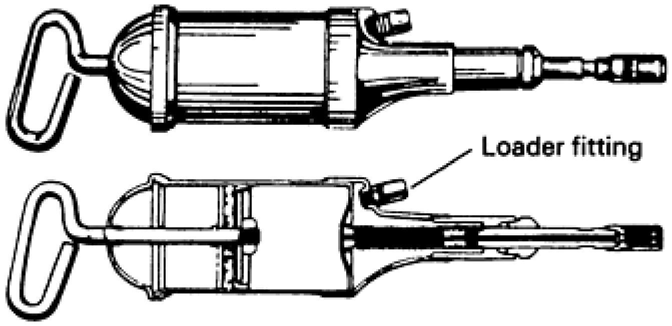
The inner construction of all hydraulic fittings includes a spring holding a steel ball against the grease-entrance hole. Grease that enters under pressure forces the ball back against the spring, but, when the pressure is stopped, the spring forces the ball back into place, effectively sealing the hole against the escape of grease or entrance of dirt.

Grease fittings must be wiped clean with a cloth (not waste, which shreds) before applying a grease gun. Otherwise, dirt might be forced into the fitting and carried onto the bearing. If a fitting is damaged or will not take grease when pressure is applied, it must be replaced.

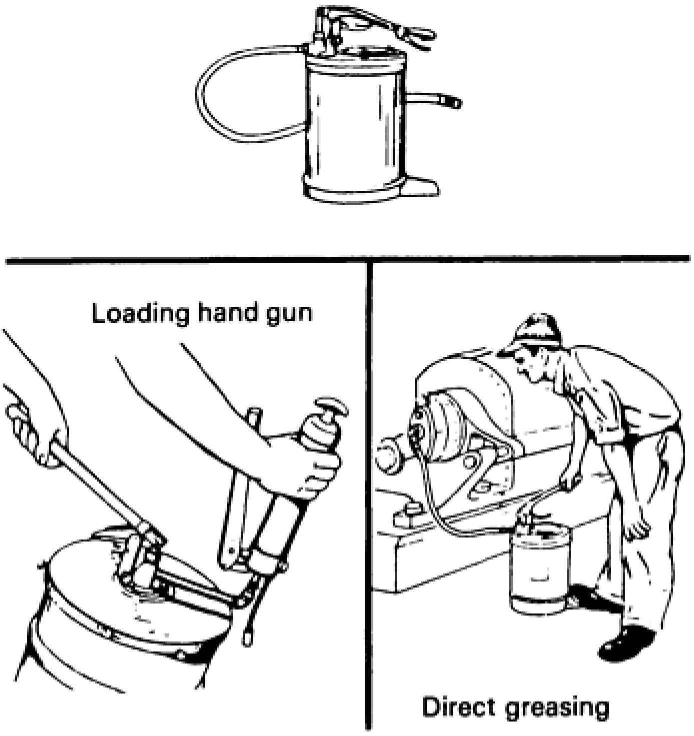
Grease Guns. Two types of hand-operated grease guns that hold from 0.5 to 1.5 kg (1 to 3 lb) of grease are commonly used. The lever gun is used primarily to grease large pieces of machinery where a large volume of grease is needed and pressure up to 70 MPa (10 ksi) is desirable. The gun is filled through the loader fitting at the front. Grease is then forced into the fitting by pumping the lever. As the grease is ejected, the spring pushes the plate forward, which keeps the remaining grease packed and in position for use.



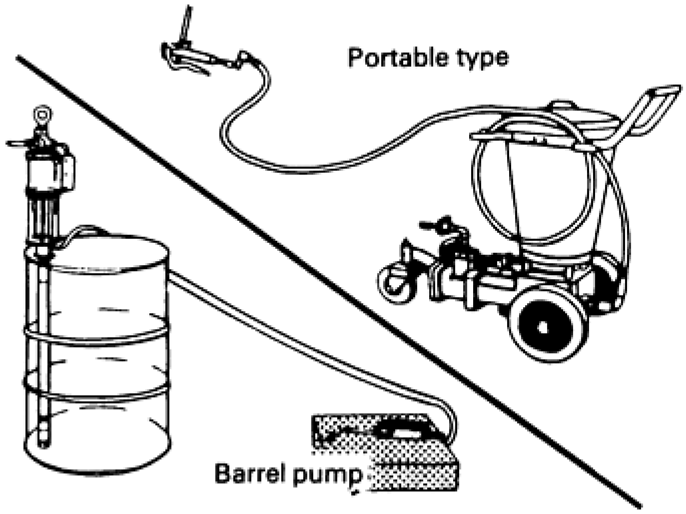
The push gun is generally used for delivering smaller quantities of grease at lower pressures to electric motor bearings and smaller machinery in the steel plant. It is filled through a loader fitting in the front. To force grease into the bearing, the nozzle is placed against the grease fitting and the handle is pushed, causing a pumping unit in the front of the gun to force grease into the bearing.



A bucket pump, which can hold about 16 kg (35 lb) of grease, is used when a greater grease-holding capacity is desired. It can be used to either fill hand guns through loader fittings or to pump grease directly into bearings through grease fittings. The bucket pump is placed on the floor and the handle is worked to deliver grease through a hose to the bearing. The bearing is usually equipped with a buttonhead fitting if it is to be lubricated with a bucket pump. The figure below illustrates a grease gun being filled by a portable loaded pump placed in a drum of grease.



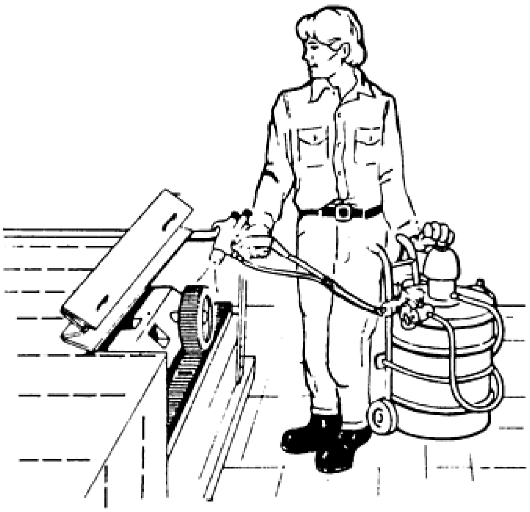
Air-operated grease pumps can be used to advantage when large numbers of bearings require a considerable volume of grease. A typical air-operated grease pump has an 18 kg (40 lb) grease reservoir mounted on wheels. Before the pump can be operated, it must be connected by hose to a compressed-air supply. Grease is delivered through a hose and hand-control valve to the bearings being lubricated.



Another type of power pump is the air-operated barrel pump, which is inserted directly into a grease drum. Barrel pumps can be used to pump grease directly to bearings, portable grease pumps and buckets, or through centralized grease systems.

It should be mentioned at this point that mobile-power grease guns, operated by electricity, are convenient to use at points beyond the range of the plant compressed-air lines.

Spray guns represent another grease application method. The guns, which are similar to those that deliver paint, use compressed air to spray gear teeth with a film of grease sufficient for lubrication. Grease is delivered to the spray gun through a hose from an air-operated barrel pump inserted in the grease drum. This method is not as wasteful as hand application, which tends to apply an excess of grease that is eventually squeezed out and lost.



Centralized grease systems supply a number of points of use from one central reservoir. If properly serviced and maintained, this system is the best method for supplying grease to a large number of bearings on a machine. Such a system consists of a centrally located grease reservoir with a pump and permanently installed piping with grease distribution valves, through which grease is conveyed from the reservoir to the various bearings. When the pump is operated, each bearing receives a definite quantity of grease, thereby eliminating the need to grease each bearing individually by hand. The pump can be operated by hand, by compressed air, or by an electric motor. If it is air or motor operated, the system can be equipped with a timer, which automatically starts and stops the pump at predetermined intervals.

Frequency (Ref 14)

The period during which a grease-lubricated bearing will function satisfactorily without relubrication is dependent on the bearing type, size, speed, operating temperature, and the grease used. The relubrication intervals (hours of operation) obtained from Fig. 8 are valid for bearings in stationary machines where loading conditions are normal. The diagram is based on the use of an age-resistant, average-quality grease and is only valid for bearing operating temperatures up to 70 °C (158 °F), measured at the outer ring. The intervals should be halved for every 15 °C (27 °F) increase in temperature above 70 °C (158 °F), but the maximum permissible operating temperature for the grease must not be exceeded. However, it should be noted that relubrication intervals can vary significantly even where apparently similar greases are used.

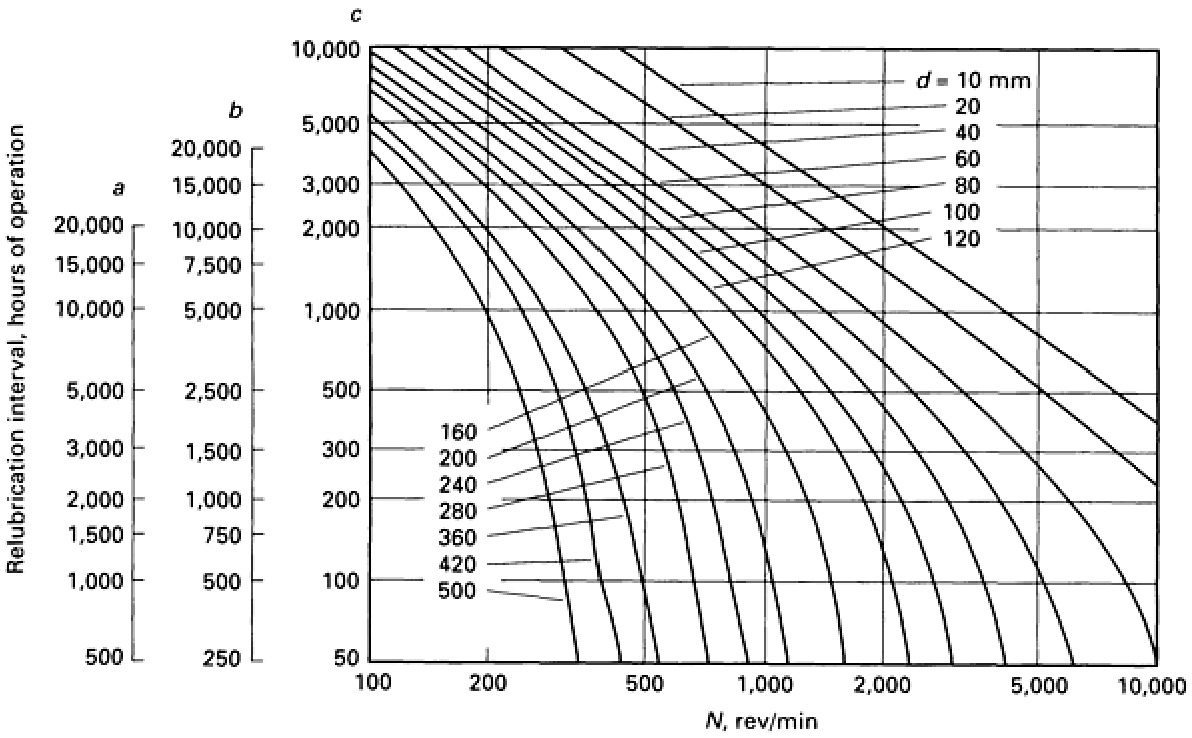


Fig. 8 Grease relubrication intervals. *a,* radial ball bearings; *b,* cylindrical or needle roller bearings; *c,* spherical or taper roller bearings and thrust ball bearings; *d,* bearing bore diameter. Courtesy of SKF-USA, Inc.

For small bearings, particularly deep-groove ball bearings, the relubrication interval is often longer than the life of the bearing application and relubrication is not normally required. Ball bearings fitted with shields or seals that are "lubricated for life" are a viable alternative in such cases.

Where there is a definite risk of the grease becoming contaminated, the relubrication intervals should be reduced. This reduction also applies to applications where the grease is required to seal against moisture. For example, bearings in paper-making machines (where water runs over the bearing housings) should be relubricated once a week.

Overgreasing/Undergreasing

Overgreasing is one of the biggest reasons for bearing failures. Undergreasing is rare. Heat is the principal enemy of a bearing, and overgreasing causes the bearings to run at higher temperatures, sometimes to the point of smoking. The reason for the high temperature is the inability of the balls or rollers to move if packed full with grease. Normally, a bearing should be from one-quarter to one-third full. The bearing manufacturer can advise on the proper amount of grease to lubricate a specific bearing. This amount of grease can be translated into so many shots of grease from a grease gun or other grease application method. In critical bearing areas, a grease-quantity apparatus can be purchased.

Storage and Handling

Storage and handling of grease is important from the standpoints of product quality assurance, housekeeping, safety, and cost. Multipurpose greases should be considered wherever possible to reduce the number of greases in inventory. A multipurpose grease should have properties such as wide operating use temperature, extreme-pressure resistance, shock-loading resistance, pumpability, and resistance to any contaminants it may encounter.

Storage in bulk containers is desirable if justified by the usage rate. Reusable bulk containers are available, eliminating the need for handling drums and drum disposal. Disposable bulk containers are also gaining popularity. Portable bulk tanks can be placed either close to the area requiring lubrication or in a central storage area. If the latter is desired, it should be neat, clean, well organized, and, preferably, inside at either room or ambient temperature.

If outside storage in drums or other containers is essential, the drums are best stored on their sides under a roof, because most grease drums have bung holes on the lid or top. Any water that stands on top of the bungs is likely to get sucked into the drum, no matter how tight the drum is sealed or capped.

References

1. "NLGI Lubricating Grease Guide," National Lubricating Grease Institute, 1987, p 1.1
2. *Compilation of ASTM Standard Definitions,* 5th ed., ASTM, 1982
3. C.J. Boner, "Manufacture and Applications of Lubricating Greases," National Lubricating Grease Institute, p 1
4. T.C. Wilson, "Grease--Its Creation and Destruction," presented at American Society of Lubrication Engineers (Chicago), 4-6 May 1970, p 2
5. "Glossary of Grease Terms," Witco, Inc.
6. R.C. Gunderson and A.W. Hart, *Synthetic Lubricants,* Reinhold Publishing, 1962, p 30-31
7. M.J. Vold and R.D. Vold, *J. Inst. Petroleum Tech.,* Vol 38, 1952, p 155-163
8. "Short Course in Lubrication," Cato Oil & Grease Co., p 16, 26
9. *Petroleum Products and Lubricants,* Vol 05.1, 05.02, and 05.03, *Annual Book of ASTM Standards,* ASTM
10. "The Lubrication Engineers Manual," United States Steel, available from American Iron & Steel Engineers
11. J. Root, Witco, Inc., personal communication
12. *Principles of Lubrication,* Lubetext DG-5A, Exxo, 1973, p 14
13. E.H. Myers, "Incompatibility of Greases," presented at the NLGI Annual Meeting, Oct 1982; published in *NLGI Spokesman,* April 1983, p 24-28
14. "Bearing Installation and Maintenance Guide," SKF-USA, Inc. brochure, 140-710, 50M 8/88 CW, Aug 1988, p 38-39

Lubricants for Rolling-Element Bearings

John Rumierz, SKF USA, Inc.

Introduction

PROPER LUBRICATION of ball and roller bearings is essential to their successful performance. Guidelines for the lubrication of rolling-element bearings differ markedly from the practices established for plain or sleeve-type bearings. The science behind their lubrication is fundamental for the correct design of bearing installations and for the institution of maintenance procedures necessary to meet full service lifetimes.

Lubricants perform several functions in their role to prevent premature failure of rolling element bearings. These functions include:

- Lubrication of the sliding contact between the retainer and other bearing components
- Elastohydrodynamic lubrication of all true rolling contacts
- Lubrication in that part of the contact between the raceways and the rolling elements not represented by true rolling

Other functions of the lubricant, especially under harsh environmental conditions, are also significant. It is often required to minimize functional power losses in a bearing, and it must protect the internal surfaces of the bearing from chemical attack and corrosion. Internally, a flushing function is important, serving to remove wear debris from the rolling contact path; externally, a sealing function is significant, preventing extraneous dirt from entering the roller contact paths. The

lubricant has a strong influence on frictional noise level performance and can provide a damping medium for vibration. It will effect the thermal dynamics of a bearing, acting as a heat transfer medium to cool the bearing and to redistribute the internal heat in order to minimize differential thermal expansions.

Obviously, no specific product or class of lubricating products can meet the wide variety of rolling bearing application requirements. Operating conditions that range from cryogenic to ultrahigh temperature, or from partial rotation to very high speeds, and involve a variety of hostile chemical environments all must be accommodated. A compromise is often necessary between performance levels and cost considerations. The latter involves not only initial lubricant costs and application system expense, but also the effect of bearing lifetime and replacement on the cost of the entire mechanical system.

Liquid Lubricants

Liquid lubricants for rolling bearing applications are usually refined mineral oils. However, the wide variation in petroleum-base stocks, with respect to molecular species and refining techniques, gives rise to oils of differing physical and chemical properties. Essentially all oils are additive-enhanced for both viscosity retention and performance improvement.

Synthetic hydrocarbon fluids (SHC) are manufactured from petroleum-base stocks. These products are synthesized with specifically chosen molecular compounds to provide favorable physical properties for lubrication. Because of their chemical similarity, SHC fluids are miscible and compatible with refined mineral oils.

Other synthetic fluids are manufactured from nonpetroleum-base chemicals. Included in this class are dibasic acid esters, phosphate esters, silicones, silicate esters, and fluorinated polyalkylethers, none of which are compatible with petroleum-base lubricants.

The advantages of fluid lubrication, versus other types of lubrication, are:

- In applications with high contamination levels, fluids are easier to remove and refill
- Lubricant quantity is more easily and accurately controlled, depending on the application
- Fluids may be more suitable for lubricating the multicomponent system
- When used with a cooler, a circulating fluid system permits the lubrication of bearings at temperatures well beyond the oxidative capacity of the bulk fluid

Fluid Lubrication for Rolling Bearings

For most applications that do not involve operating conditions of extremely high temperatures, good-quality refined mineral oils are satisfactory. Under extreme operating conditions, one of the synthetic fluids may be preferred. In either case, the lubricant must be free of contamination that can cause wear or corrosion in the bearing. Furthermore, the fluid should resist evaporation, oxidation, and viscosity change under the conditions of operation.

Internal heat generation, friction torque, and, eventually, bearing lifetime are all affected by two main factors. One is dependent on the bearing design and loading particulars and the other is a function of lubricant quantity, viscosity, and the bearing speed. Only enough lubricant is required to permit formation of a thin film between contacting surfaces. Both too much lubricant and too high a fluid viscosity will adversely affect friction and torque.

Heat generation in a bearing is dependent on torque and speed. Lubricant temperature rise will always result in decreased fluid viscosity and a decrease in frictional torque. The overall heat balance of the bearing and surrounding structure will determine the steady-state condition.

No one lubricant is suitable for all bearing applications. When operation calls for a wide temperature range, a fluid with a high viscosity index, that is, the least change in viscosity versus temperature, should be chosen. Applications with very low starting temperatures require fluids with a sufficiently low pour point for free bearing rotation upon low-temperature start-up. The bearing manufacturer should be contacted for specific recommendations when unusual conditions exist.

Methods. Proper enclosures and supply systems must be provided to contain fluid lubricants and thereby prevent leakage and deterioration of fluid and the bearing. Several of the most widely used containment systems are described below.

Oil bath or sump lubrication provides a minimum amount of fluid in constant contact with the bearing, as shown in Fig. 1. This configuration is useful in vertical, horizontal, and inclined axis applications. In the static condition, the lubricant level just reaches the center of the lowest rolling element. Higher levels will result in churning losses, which can cause fluid oxidation and early failure. Too low a lubricant level leads to windage, lubricant starvation, and early failure. A gage must be provided to maintain the proper fluid level.

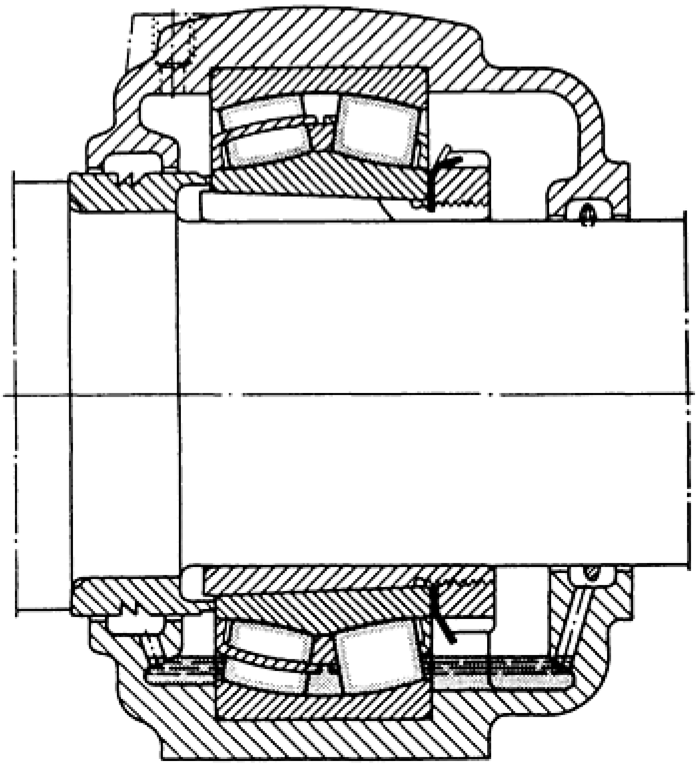


Fig. 1 Oil bath method of bearing lubrication

Oil bath systems are widely employed in hot applications where grease purging cannot be tolerated. Fluid circulation provides better heat dissipation and cooler running temperatures. This aspect is important under conditions of heavy load, where contact friction losses dominate viscous churning losses. The use of a heat pipe or cooling coils with an oil bath system can greatly extend the application temperature of the fluid.

Shaft motion can be used to drive a viscous pump for fluid supply, reducing the sensitivity of the system to the fluid level. A disk drags oil from the sump up a housing groove to a scraper blade that deflects the oil to a drilled passage leading to the bearing. A limitation of sump systems is the lack of effective filtration. Such systems are only suitable for inherently clean-running applications.

Circulating Fluid Systems. Operation at higher speeds will cause the temperature to increase and will accelerate oxidation of the lubricant. In such situations, a circulating system provides significant advantages over static systems. Pumping a fluid lubricant through the bearing into a reservoir increases the amount of heat dissipation. A pressure feed arrangement allows the use of heat exchangers to either remove or add heat as the operating conditions dictate, as shown in Fig. 2.

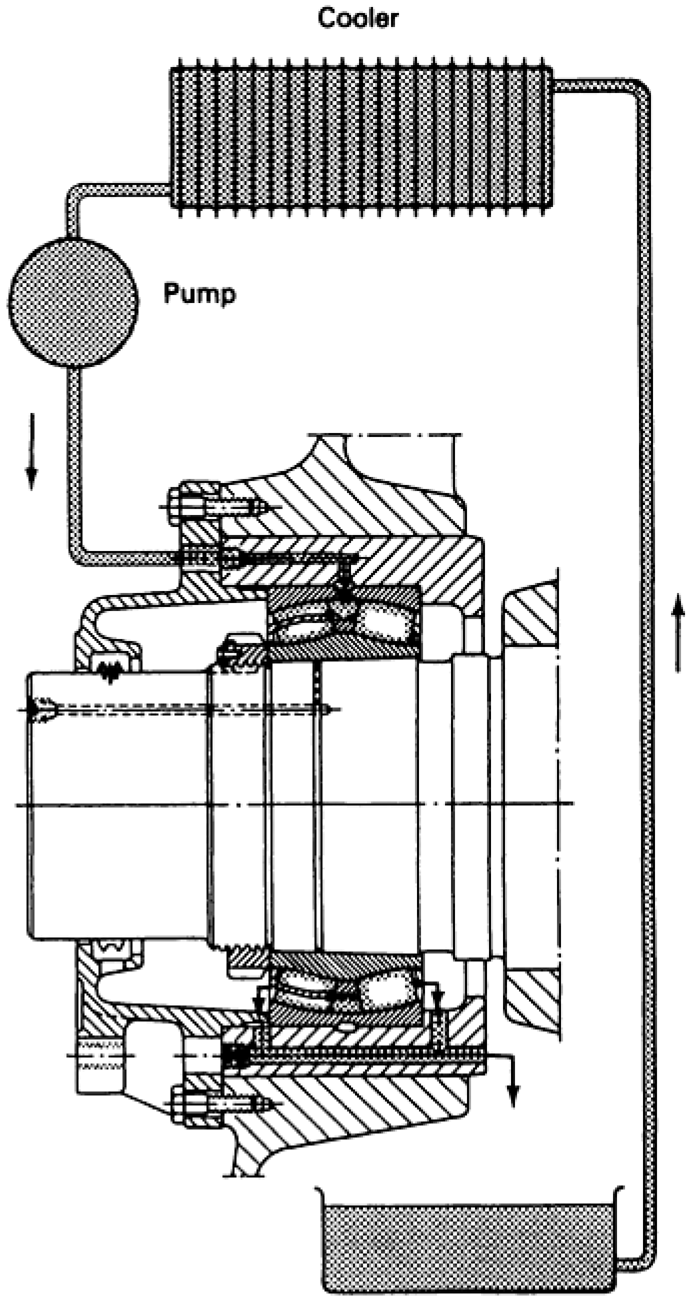


Fig. 2 Circulating oil lubrication system

Circulating systems are usually fitted with filtration systems to remove bearing wear and environmental contamination from the lubricant. The use and maintenance of clean oil is of utmost significance in obtaining extended bearing life. Generally, increased filtration is beneficial to bearing lifetime if the filter is properly chosen and maintained. A 10 μm (0.4 mil) filter is a good choice. The sump should be of adequate capacity to contain oil for at least 30 min in order to provide settling of contaminants.

Oil Jet. The amount of oil for very high speed operation should be minimized to limit bearing temperature rise. A useful method for this case is oil jet lubrication. As shown in Fig. 3, a jet of oil under pressure is directed at the side of the bearing, between the inner ring outside diameter and cage inside diameter. The oil jet velocity must be high enough (>15 m/s, or 50 ft/s) to penetrate the turbulence generated by the bearing rotation.

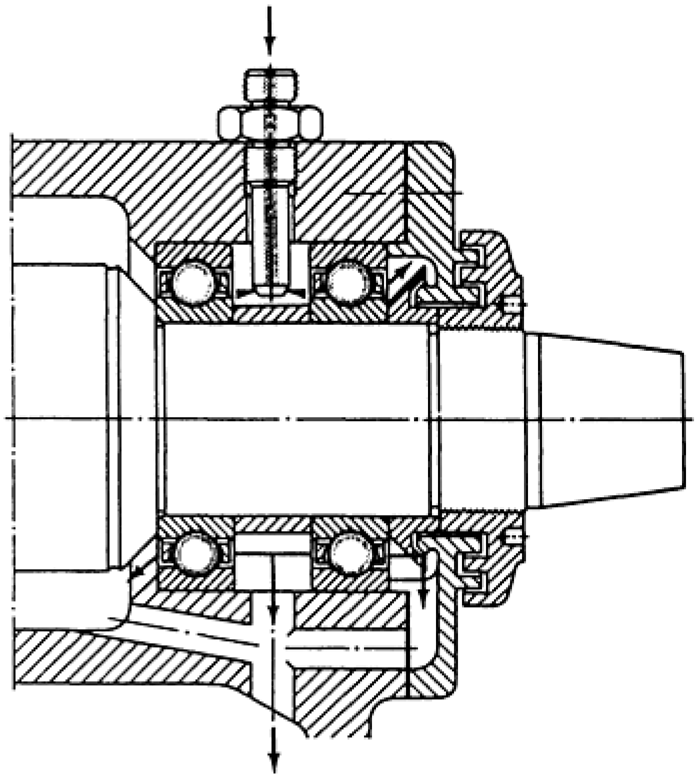


Fig. 3 Oil jet lubrication of rolling bearings

Oil mist lubrication has been found to be useful in either high-speed applications or those in which the bearing housing is surrounded by water and grit. Mist lubrication is suitable for both horizontal and vertical shafts. Very small, accurately metered amounts of oil are directed at the rolling element by compressed air. By minimizing oil quantity, operation can occur at a lower temperature and higher speed than is possible using any other method of lubrication. Oil consumption is consequently low, and the air flow prevents the entrance of grit and subsequent wear. Several atomizer designs are available from a variety of manufacturers.

Single-pass lubrication systems can be effective when minimum bearing friction is essential, loads are low, and speeds are moderate. A minimum spray of fluid mist is delivered to the bearing contacts. Churning and resultant losses are eliminated, and low usage permits discarding the lubricant after a single pass. One-time exposure to the contact environment relaxes the oxidation and property requirements of the lubricant. Tests have shown that trace quantities of injected lubricant at 1 h intervals are sufficient to keep precision spindle assemblies running at friction torque levels that are unobtainable by any other method. A disadvantage is that the generation of oil mists outside the bearing enclosure must be strictly limited to meet health and safety regulations.

Mineral Oils

The predominant chemistry for rolling bearing lubricants is refined mineral oil, which generically refers to a product of petroleum crude. Chemically, these oils are composed of a large number of paraffinic, naphthenic, and aromatic groups, combined into many distinct molecules. Also present are trace amounts of molecules containing sulfur, oxygen, or nitrogen. On an elemental basis, the composition of petroleum oils is consistent: carbon, 83 to 87%; hydrogen, 11 to 14%; and a remainder composed primarily of sulfur, nitrogen, and oxygen. The exact molecular makeup of a petroleum-base stock is very complex and is dependent on its specific origin.

With regards to lubrication, petroleum-base stocks are characterized by the chemistry of the distillates obtained. Therefore, it is common to speak of paraffinic, naphthenic, and mixed crude oils. Aromatics seldom predominate in lubricant oils. Modern distillation, refining, and blending techniques allow the production of many oil products from a given base stock. However, because of subtle variations, some base stocks are more desirable for lubricant formulation, as discussed below.

Paraffinic-base stocks have a favorable viscosity-temperature relation for lubrication. Such fluids are generally low in undesirable trace components. Naphthenic-base stocks do not contain paraffinic waxes and are better suited for certain low-temperature uses. These base stocks also have lower flash points and are more volatile, compared with paraffinic oils.

Two common industrial lubricants are the rust and oxidation (R&O) inhibited and the extreme-pressure (EP) oils. R&O inhibited oils are often formulated with additional additives, such as antifoam and antiwear agents. These products are generally suitable for use between -20 and 120 °C (-4 and 248 °F). They are often employed in applications where bearings and gears share a common lubricant reservoir.

Extreme-pressure oils are usually R&O inhibited products with an EP additive to generate a lubricating surface that can prevent metal-to-metal contact when the fluid film fails. Two main strategies exist in formulating EP oils. One is to use active sulfur, chlorine, or phosphorus compounds to generate sacrificial surfaces at the contacts. These surfaces will then shear, rather than weld, upon contact. The second approach employs a planar solid to impose between the contact surfaces. Of course, both contacting situations occur only when there is insufficient fluid film to separate them.

EP oils are used either when bearing loading is high or where shock loading exists. Normally, EP oils are used between -20 and 120 °C (-4 and 248 °F). Some cautions are necessary when such products are used. EP solids can reduce internal bearing clearances, causing wear and, possibly, failure in certain bearing types. These additives can also be lost upon filtration. Some sulfur-chlorine-phosphorus additives are corrosive to bronze and nylon cages and accessory items.

Synthetic hydrocarbon fluids are manufactured from chemical precursors, rather than the petroleum-base stocks that constitute mineral oils. Whereas a large number of molecules exist in mineral oils, the number and type of molecules in synthetic hydrocarbons are strictly controlled by the manufacturing process involved. The ability to pick and choose components allows the production of a petroleum fluid with optimum properties for lubrication. One commercially important type is the polyalphaolefin (PAO) fluids that are widely used in turbine lubricants, hydraulic fluids, and grease formulations.

PAO fluids show very high viscosity indexes, compared with refined mineral oils, which means better viscosity retention at elevated temperatures. Synthetic hydrocarbon lubricants exhibit superior thermal and oxidation stability over conventional lubricants, permitting higher operating temperatures. Other improved properties include flash point, pour point, and volatility characteristics. Although synthetic, the materials are compatible with refined petroleum lubricants because of the similar chemistry involved.

Viscosity of Lubricants

The most important property of a lubricant under normal conditions is viscosity. This applies both to fluid lubricants and to the base fluids in grease formulations. By definition, viscosity is the resistance to flow. For the purposes of this article, viscosity is the factor of proportionality between shearing stress and the rate of shear. Very simply, increased viscosity relates to an increased ability of the lubricant to separate contacting microsurfaces under pressure. This separation is at the heart of lubrication for rolling bearings.

Viscosity is usually measured kinematically per ASTM D 445. This test measures the time required for a measured volume of fluid to pass through a standard length of capillary tube under the force of gravity. Standardized test temperatures for rolling bearing lubricants are between 40 and 100 °C (104 and 212 °F). Many alternative viscosity determinations exist and are of utility when either very viscous fluids or low temperatures are involved.

The ISO VG classification is universally used to designate lubricant viscosity grade. This classification is based on the ISO 3448-1975(E) standard. Simply put, an ISO VG 32 lubricant has an approximate viscosity of 32 mm^2/s (32 cSt) at 40 °C (104 °F). A range of viscosities is defined for each grade in the ISO standard. ISO grades run from VG 2 to VG 1500.

The derived quantity, viscosity index (VI), is often encountered. This dimensionless number reflects the effect of temperature on kinematic viscosity. The higher the VI for a fluid, the smaller the viscosity loss with increased temperature. A typical paraffinic mineral oil base lubricant will have a VI from 85 to 95. Polymers can be added to mineral oil base stocks to obtain a VI of 190 or more. The shear stability of these additions, as well as the actual effect in the microcontact, is open to question, and the VI of such fluids generally deteriorates with time. Many of the synthetic fluids have VIs that far exceed those of mineral oils. ASTM D 567 describes the method of calculating VI from kinematic viscosities at two temperatures.

Selection of Proper Viscosity for Petroleum Oil Lubricants. Figures 4 and 5 can be used to obtain a minimum acceptable viscosity for a bearing application. With known values for the pitch diameter, d_m, and the rotational speed, Fig. 4 can indicate the minimum suitable viscosity at the bearing operating temperature. Figure 5 can then be used to relate this viscosity to the standard reference viscosities for ease of selection. Figure 5 can also be used to determine the actual viscosity of a petroleum oil with a VI of 85 at a given temperature if its standard viscosity data are known.

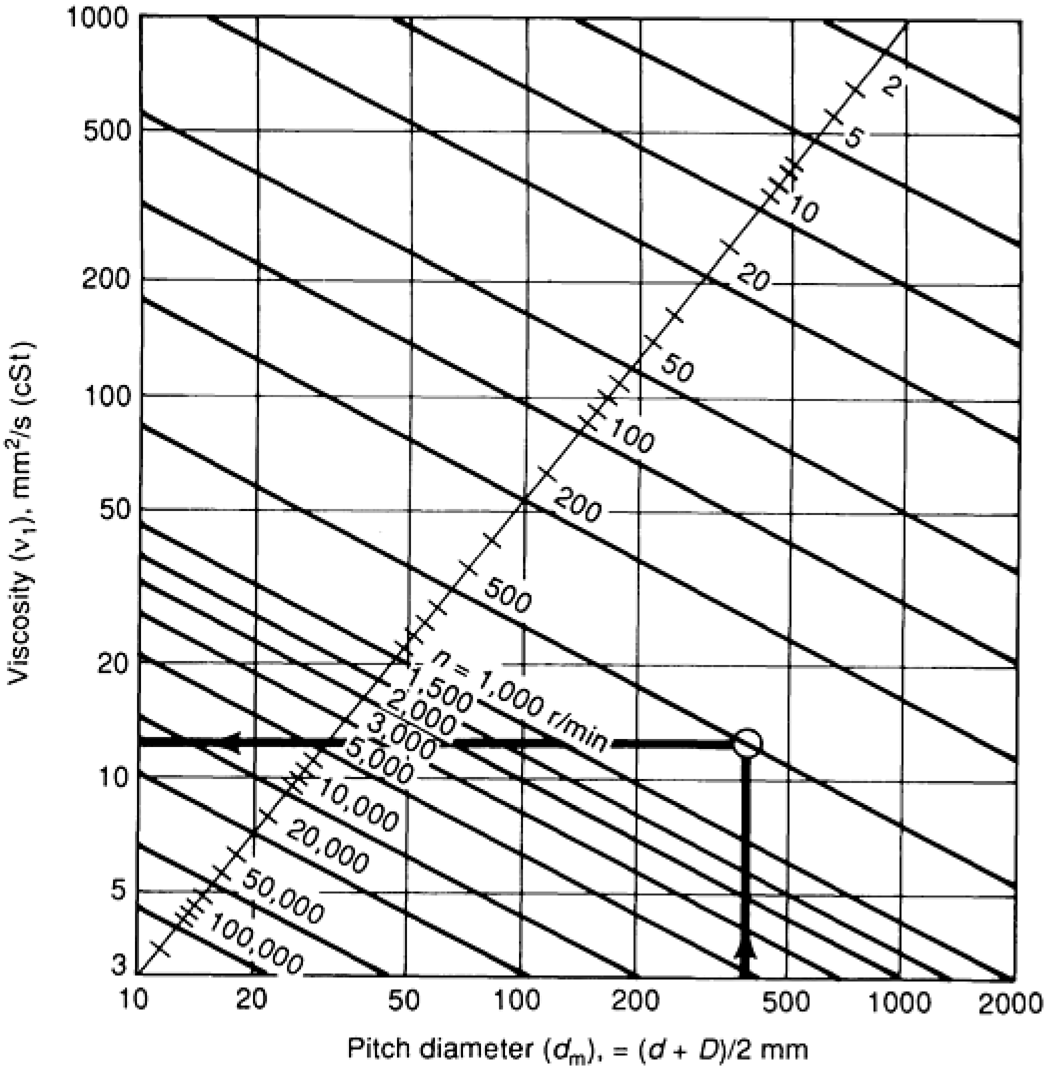


Fig. 4 Calculation of minimum required viscosity

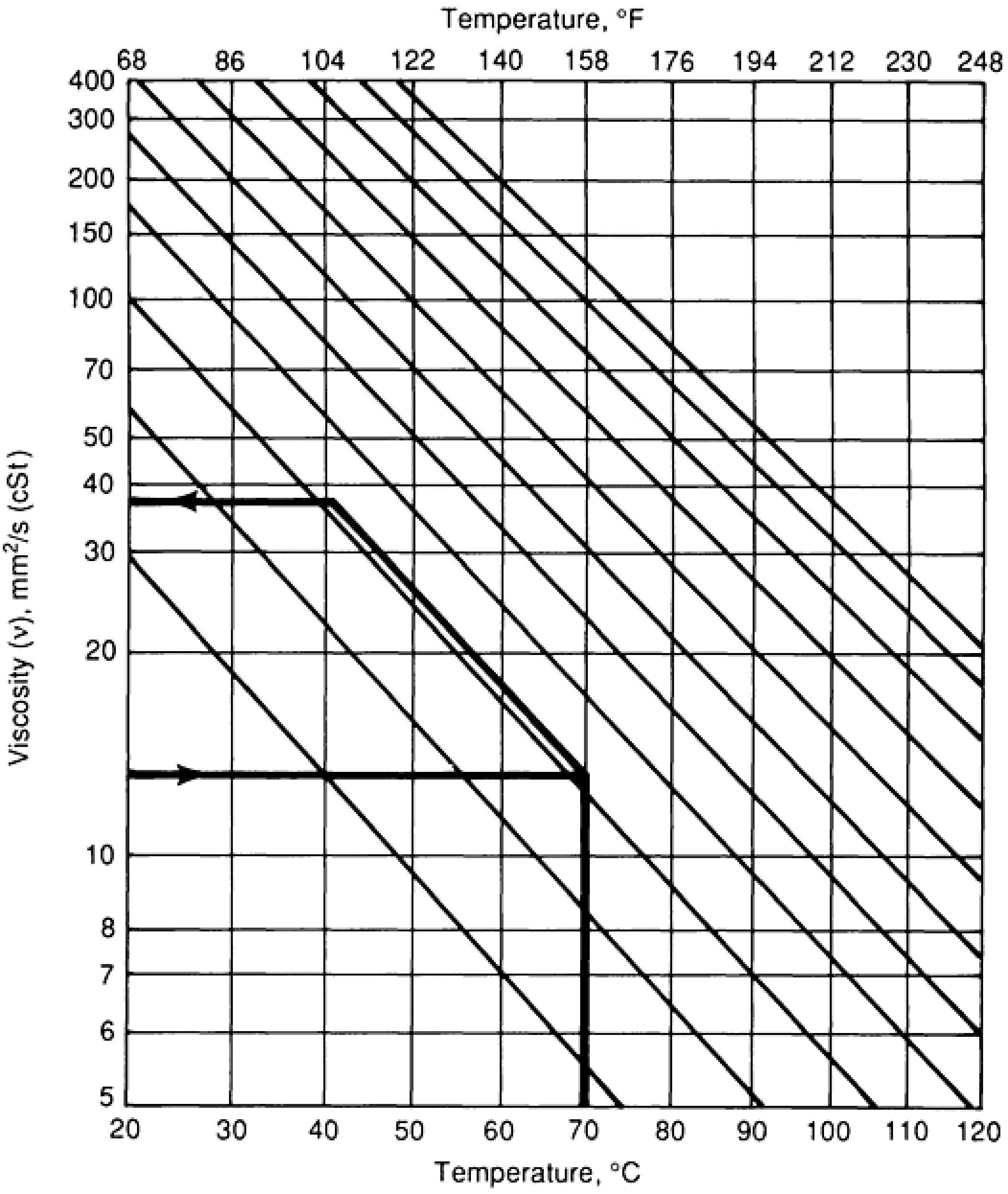


Fig. 5 Viscosity-temperature relation for mineral oil base lubricants with a VI of 85

Example of Viscosity Calculation. A bearing has a bore diameter, *d,* of 340 mm (13.6 in.) and an outside diameter, *D,* of 420 mm (16.8 in.). Thus, its pitch diameter is 380 mm (15.2 in.). It is operating at 70 °C (160 °F) and at 500 rev/min. What is the minimum acceptable viscosity under these conditions? As shown in Fig. 4, the required kinematic viscosity is at least 13 mm^2/s (13 sSt). Remembering that the operating temperature is 70 °C (160 °F), it can be seen, in Fig. 5, that the required viscosity of an oil at 40 °C (104 °F) is at least 39 mm^2/s (39 cSt).

When estimating operating temperature, it is useful to remember that the oil temperature is from 3 to 11 °C (5 to 20 °F) higher than the bearing housing temperature. If a lubricant with a higher than required viscosity is used, an improvement in bearing fatigue life is expected. However, because increased viscosity will raise the operational temperature, there is a practical limit to the lubrication improvement that can be obtained by these means. When unusually high or low speeds, or heavily loaded conditions, or unusual lubrication circumstances are encountered, the bearing manufacturer should be consulted.

Types and Properties of Nonpetroleum Oils

Many types of "synthetic" fluids have been developed in response to lubrication requirements that are not adequately met by petroleum oils. These requirements include extreme temperature, fire resistance, low viscosity, and high viscosity index.

Table 1 lists typical properties of various lubricant base stocks and indicates application areas for finished products of each type. As is the case for petroleum oils, many additive chemistries have been developed to enhance the properties of these fluids.

Table 1 Typical properties of lubricant base fluids

Base fluid	Density, g/cm^3	Viscosity at 40 °C (104 °F), mm^2/s	Viscosity index	Flash point		Pour point		Volatility(a)	Oxidation resistance(a)	Lubricity(a)	Thermal stability(a)	Application range
				°C	°F	°C	°F					
Mineral oils	. . .	. . .	. . .	. . .	. . .	. . .	. . .	5	5	5	5	Standard lubricant components, either as oil or grease base
Paraffinic	0.881	95	100	210	410	-7	19	. . .	. . .	. . .	. . .	
Naphthenic	0.894	70	65	180	356	-18	0	. . .	. . .	. . .	. . .	
Mixed base	0.884	80	99	218	424	-12	10	. . .	. . .	. . .	. . .	
VI improved	0.831	33	242	127	261	-40	-40	. . .	. . .	. . .	. . .	
Polyalphaolefin	0.853	32	135	227	441	-54	-65	4	4	5	4	
Esters	. . .	. . .	. . .	. . .	. . .	. . .	. . .	3-5	3-7	3-6	4-7	Jet turbine lubricants, hydraulic fluids, heat transfer products; used as bases for low-volatility, low-viscosity grease; phosphate esters fire resistant
Diabasic	0.945	14	152	232	450	<-60	<-70	3	5	5	5	
Polyol	0.971	60	132	275	527	-54	-65	. . .	. . .	. . .	. . .	
Tricresyl phosphate	1.160	37	-65	235	455	-23	-9	3	3	3	7	
Polyglycol ether	0.984	36	150	210	410	-46	-51	3-5	7	4	5	
Silicate	0.909	6.5	150	188	370	<-60	<-76	3	7	6	4	
Silicones	. . .	. . .	. . .	. . .	. . .	. . .	. . .	1	1	7	1-5	High-temperature/low-volatility applications, either as oil or grease base; lightly loaded bearings; excellent thermal stability
Dimethyl	0.968	100	400	>300	>572	<-60	<-76	. . .	. . .	. . .	. . .	
Phenyl methyl	0.990	75	350	260	500	<-60	<-76	. . .	. . .	. . .	. . .	
Chlorophenyl	1.050	55	160	288	550	<-60	<-76	. . .	. . .	. . .	. . .	
Perfluoroethyl	1.230	44	158	>300	>572	<-60	<-76	. . .	. . .	. . .	. . .	
Phenyl ether	. . .	. . .	. . .	. . .	. . .	. . .	. . .	1	1-3	5-7	1	
Low-viscosity	1.180	75	-20	263	505	10	50	. . .	. . .	. . .	. . .	
High-viscosity	1.210	355	-74	343	650	5	41	. . .	. . .	. . .	. . .	
Perfluoroalkylether	1.910	320	138	. . .	. . .	-	-	3-7	1	5-7	1-3	Extreme-temperature fluid; used in very low

	32	26	volatility applications

(a) The value 5 characterizes highly refined mineral oil. Values of <5 reflect superior performance, whereas values >5 reflect inferior performance to mineral oil with respect to lubrication properties.

Before discussing the general synthetic classes, it should be noted that the use of such lubricants requires a thorough understanding of the application requirements. The favorable properties of some synthetics are obtained only with unsuitable performance characteristics in areas such as load-carrying ability and high-speed operation. Similarly, many very high temperature fluids developed for military applications have very short service lifetimes, compared with commercial requirements.

Polyglycols are often used as a synthetic lubricant base in water emulsion fluids. This class of fluids includes glycols, polyethers, and polyalkylene glycols. Properties of the class include excellent hydrolytic stability, high viscosity index, and low volatility. The most prevalent usage is as a component of fire-resistant hydraulic fluids.

Phosphate esters have poor hydraulic stability and a low viscosity index. Because an outstanding characteristic of these fluids is fire resistance, they are often used as hydraulic fluids in high-temperature applications, such as aerospace.

Dibasic acid esters are a family of synthetic base stocks that are widely used in aircraft turbine applications and as a basis for low-volatility lubricants. They are synthesized by reacting aliphatic dicarboxylic acids (adepic to sebacic) with primary branched alcohols (butyl to octyl). Some are available from natural sources, such as castor beans and animal tallow. Characteristic properties of these fluids are low volatility and high viscosity index. Polyol esters that are formed by linking dibasic acids through a polyglycol center are suitable as high film strength lubricants.

Blends of dibasic esters, complex esters with suitable antiwear additives, VI improves, and antioxidants are used to form the current generation of jet engine lubricants. Generally, these products show excellent viscosity-temperature relationships, good low-temperature properties, and acceptable hydrolysis resistance. Elastomeric seals used with these materials must be chosen carefully, because many standard rubbers will suffer attack.

Silicone fluids (organosiloxanes) exhibit outstanding viscosity retention with elevated temperature and are functional under conditions of extreme heat and cold. These fluids are the basis for many high-temperature (200 °C, or 392 °F) lubricants.

In addition to favorable viscosity-temperature characteristics, volatility is low and both thermal and oxidation resistance characteristics are excellent. As a family, these fluids exhibit good hydrolytic stability. If very high temperatures are avoided, these fluids are inert with most elastomers and polymers. However, oxygen exposure with high temperature can result in gelation and loss of fluidity.

The lubrication properties of these fluids are not impressive when compared with other classes of lubricating fluids. Typical applications are in electric motors, brake fluids, oven preheater fans, plastic bearings, and electrical insulation.

Silicate esters represent a mating of the previous two lubricant fluid types. As a class, these fluids possess good thermal stability and low volatility. They are used in high-temperature hydraulic fluids and low-volatility greases.

Fluorinated polyethers are the highest-temperature lubricating fluids commercially available. Although distinct chemical versions are marketed, all of these fluids are fully fluorinated and completely free of hydrogen. This structural characteristic makes them inert to most chemical reactions, nonflammable, and extremely oxidation resistant. Products from these oils show very low volatility and excellent resistance to radiation-induced polymerization. The products are essentially insoluble to common solvents, acids, and bases. The density of these oils is approximately double that of conventional petroleum oils.

Products of this chemical family are used to lubricate rolling bearings at extremely high temperatures from 200 to 260 °C (392 to 500 °F). Other application areas include high-vacuum operations, corrosive environments, and oxygen-handling systems. As would be expected, the cost of these oils is very high.

Grease Lubrication

Greases consist of two major components: a fluid phase and a thickener system that determines the consistency of the product. Although thickeners can comprise a variety of materials, all provide a large specific surface area to retain oil. Mineral oils predominate, but the fluids used in grease formulations encompass the spectrum of lubricating fluids. The resulting product behaves like a semisolid, releasing oil at a controlled rate to meet the requirements of the rolling bearing. The fluid phase of the grease is either gradually degraded by oxidation or lost by evaporation, centrifugal force, and other factors. In time, the grease at the contacting bearing surfaces becomes depleted.

For many applications, greases have several advantages that are not offered by fluid lubricants:

- Fluid level is no longer a consideration, which reduces maintenance
- Enclosures around the bearing can be simplified, because grease is confined to the bearing interior
- Leakage is better controlled, which minimizes contamination of foodstuffs, textiles, and chemical products
- Grease usage can provide improved bearing sealing in conjunction with properly designed seals
- After start-up, friction torque and the associated temperature rise are usually more favorable

Excess grease will cause the bearing temperature to rise, especially at high speeds. Recommended grease quantity should be determined by the bearing vendor and followed. A fuller bearing is not necessarily a better bearing, where grease is concerned. For a typical application, a grease fill that ranges from 33 to 50% of the internal bearing volume will ensure adequate lubrication. Greater amounts will be purged past the seals and wasted.

Grease Speed Limits. A maximum operating speed based on type and design specifics is set by bearing manufacturers for their products. Speed limits for grease lubrication are lower than for fluid lubrication because of the initial temperature peak that occurs at bearing start-up. Operating temperatures will drop to lower levels once the grease has been distributed through the bearing. Certain bearing designs, such as angular-contact ball bearings, full-complement cylindrical rolling bearings, and taper rolling bearings, accentuate grease churning and make even lower speed limits necessary.

Operating speed is influenced by the shear strength of the grease, which is primarily determined by thickener type. A speed factor, nd_m, is often quoted by grease manufactures to indicate the speed capability, where n is the operating speed and d_m is the mean bearing diameter in millimeters. Among various grease compositions, these values ranges from 5×10^4 to 1.5×10^6. Greases that run above their nd_m value will cause temperature excursions in the bearing.

Lubricating Grease Composition. Rolling bearing greases are either thickened mineral oils or synthetic fluids. The thickeners are often metallic soaps. Additives are used to enhance certain properties of the grease. The maximum operating temperature is dependent on the chemistry of the constituents. Grease consistency is determined by the type and concentration of the thickener. When selecting a grease for a rolling bearing application, the viscosity of the base fluid, the operating temperature range, the rust-inhibiting properties, and the load-carrying capability are usually the most significant factors.

Base Fluid Viscosity. The efficiency of lubrication is primarily determined by the degree of separation at the microcontacts of the bearing. If an adequate load-carrying fluid film is to be formed, the lubricant must have a given minimum viscosity at the operating temperature. Base fluid viscosity for rolling bearings normally falls between 15 and 500 mm^2/s at 40 °C (104 °F). Grease that are based on heavier fluids may bleed oil so slowly that the bearing surfaces will not be adequately lubricated. Greases with very high viscosity fluids are recommended in only very special circumstances. Generally, oil lubrication is more reliable for high-viscosity requirements. Base fluid viscosity also governs the maximum operation speed of a grease in a given application. For very high speeds, the most suitable greases will have base fluids of very low viscosity.

Consistency of Lubricating Greases. Consistency refers to how stiff or soft a grease is. Stiffer products offer better sealing, whereas softer greases show less resistance to flow. Greases are categorized by consistency classes according to the National Lubricating Grease Institute (NLGI) scale (Table 2). ASTM D 217 is used to generate the penetration values upon which the greases are classified.

Table 2 NLGI penetration grades

NLGI grade	Penetration (60 strokes)[(a)]
000	445-475
00	400-430
0	355-385
1	310-340
2	265-295
3	220-250
4	175-205

5	130-160
6	65-115

(a) Per ASTM D 217

The consistency of rolling bearing greases should not dramatically change with temperature or with mechanical working. Greases that soften excessively, either with temperature or with working, may leak from the bearing, whereas those that stiffen at low temperature may restrict bearing rotation. Greases of NLGI grade 1, 2, or 3 are generally used for rolling bearing applications. NLGI grade 3 greases are often used for vertical shaft arrangements to prevent grease leakage.

Some greases that are thickened with polyurea can soften and harden reversibly, depending on shear rate, being relatively soft under low shear and stiff under high shear. The use of such greases should be restricted to ball bearings in horizontal applications.

Temperature Range. The operating temperature range of a rolling bearing grease depends on the chemical type of base fluid and thickener used. The lower temperature limit is dependent on the base fluid type and viscosity. Good low-temperature performance is important for applications below -20 °C (-4 °F), where lubricant stiffness results in increased starting resistance and running power requirements. The maximum temperature is determined by the thickener type and the oxidative stability of the base fluid. The upper operational temperature limit should not be confused with the dropping point value quoted by lubricant manufacturers. Dropping point indicates the temperature at which the grease loses its consistency and becomes fluid. The temperature can be well in excess of the maximum serviceable temperature of the grease. Greases based on synthetic oils, ester oils, synthetic hydrocarbons, or silicones can be used at temperatures above or below the operating temperatures of greases based on mineral oil.

Table 3 gives the operating temperature ranges for the most commonly used rolling bearing greases based on thickener type. The ranges are valid for greases that have a mineral oil base but no EP additives. Of the grease types listed, lithium 12-hydroxystearate soap thickened greases are the most commonly used for rolling bearing lubrication.

Table 3 Properties of mineral oil base grease with various thickeners

Thickener	Dropping point		Maximum temperature				Low-temperature limit		Water resistance(a)	Load-carrying capability(a)	Corrosion protection(a)
			Continuous		**Peak**						
	°C	**°F**	**°C**	**°F**	**°C**	**°F**	**°C**	**°F**			
Soap type											
Aluminum complex	>230	>446	150	302	177	350	-30	-22	G-E	G	G-E
Barium complex	>200	>392	150	302	177	350	-30	-22	E	G-E	E
Calcium	90	194	60	140	77	170	-20	-4	E	G	G-E
Calcium complex	>250	>482	140	284	177	350	-30	-22	G-E	G-E	E
Lithium	180	356	121	250	150	302	-35	-31	F-G	F-G	G-E
Lithium complex	>250	>482	150	302	177	350	-35	-31	G-E	G	E
Sodium	180	356	100	212	121	250	-20	-4	P-F	F-G	P-F
Sodium complex	240	464	121	250	150	302	-30	-22	G	G	E
Nonsoap type											
Polyurea	>225	>437	150	302	177	350	-20	-4	E	F	E
Inorganic (clay)	>250	>482	150	302	177	350	-20	-4	E	F	G-E

(a) P, poor, F, fair, G, good; E, excellent

Relubrication intervals for rolling bearings depend not only on the lubricant type and environmental conditions, but on specific design features, which vary among bearing manufacturers. Although the manufacturer should be consulted for relubrication interval recommendations for their products, some general considerations can be given. As conditions become more severe in terms of operating temperature and/or frictional heat generated, the bearing must be relubricated more frequently.

As a rule of thumb, temperatures above 70 °C (158 °F) will shorten the relubrication interval of good-quality, lithium soap greases. Beyond 70 °C (158 °F), increments of 15 °C (27 °F) will halve the interval. This is useful in estimating the performance of a grease at elevated temperature, given a fixed reference point. For example, knowing that a certain grease

gave 500 h of bearing life at 150 °C (300 °F) allows estimation of its lifetimes at 165 °C (330 °F) (250 h) and 135 °C (275 °F) (1000 h).

Bearings that operate under 70 °C (158 °F) will probably require relubrication less often. However, a doubling phenomena with 15 °C (27 °F) temperature reductions has not been shown. Bearings that operate on a vertical shaft require relubrication approximately twice as often as horizontal shaft applications.

Relubrication interval requirements will vary according to the types of grease used and may also vary among greases of similar types. For small ball bearings, the relubrication interval is often longer than the life of the bearing application and therefore relubrication is not required. Where marked contamination occurs, the recommended relubrication intervals should be reduced. This is also true in applications where moisture intrusion in significant.

Corrosion Prevention Behavior. The rust-inhibiting properties of greases are generally determined by additives to the grease formulation. A grease must provide protection against corrosion and should not be easily washed out in cases of water intrusion.

Load-Carrying Ability. Under very high loads, the lubricant film that separates the contacting surfaces may become discontinuous, causing high bearing wear and premature failure. For heavily loaded bearings, it has been customary to recommend the use of grease containing EP additives, because these additives increase the load-carrying ability of the lubricant film. Previously, lead-based compounds dominated the EP additive field, often to the benefit of greases formulated with them. Because of health and environmental concerns, lead-base additives have been superseded by other compounds, some of which have proven aggressive to bearing steels at high temperatures. Reductions in bearing lifetimes have been recorded in some instances. EP additives should be used only with the utmost care to prevent reduced bearing life.

Grease Compatibility. Mixing greases of chemically different thickeners and/or base fluids can produce an incompatibility, resulting in a loss of lubrication and bearing failure. Mixing greases of differing base fluids can result in a two-component fluid phase that will not provide a continuous lubrication film. In some cases, mixing different thickeners can result in a grease mix that is either too stiff to lubricate properly or too fluid to remain in the bearing cavity. Early failure can be expected in either situation. Lubricants of differing chemical makeup should not be mixed. If there is doubt, then new lubricant should be used to purge the bearing cavities and supply lines until all traces of the prior product are removed before starting operation.

Regreasing Procedures. If the relubrication interval specified by the manufacturer is less than approximately six months, an additional grease charge can be made directly to the bearing, in accord with the manufacturer's recommendation. For intervals greater than six months, all of the used grease should be removed from the grease cavities and replaced with fresh grease.

Polymeric Lubricants

A polymeric lubricant uses a matrix, or sponge-like material, that retains its physical shape and location in the bearing. Lubrication is provided by the oil alone after it has bled from the polymeric sponge. Although ultrahigh molecular weight polyethylene forms a pack with generally good performance properties, it is temperature limited to about 100 °C (212 °F), precluding its use in some applications within the temperature capability of standard rolling bearings. Other higher-temperature materials, such as polymethylpentene, form excellent porous structures, but are relatively expensive and suffer higher torque. Fillers and blowing agents, which are standard tools of the plastics industry, interfere with the oil flow behavior and therefore contribute little in this situation.

Despite its temperature limitations, a polyethylene-base material, designated W64 by SKF, has achieved success in the solution of everyday bearing lubrication problems. Perhaps the most notable success occurs where a bearing must operate under severe acceleration conditions, typical of those found in planetary systems. Although the bearing rotational speed about its own axis may be moderate, the centrifuging action that is due to planetary motion is strong enough to throw conventional greases out of the bearing, despite the presence of seals. When polymerically lubricated bearings are substituted, life improvements of two orders of magnitude are not uncommon. Such situations occur in cable making, tire winding, and textile mill applications.

Another major market for polymer lubricants is food processing, where machinery must be cleaned frequently, often daily, using steam, caustic, or sulfamic acid solutions. Because these degreasing fluids tend to remove lubricant from the

bearings, it is standard practice to follow every cleaning procedure with a relubrication sequence. Polymer lubricants have proven to be highly resistant to washout by such cleansing methods, which reduces the need for regreasing. After washing down, the bearings should be rotated to prevent static corrosion.

The reservior effect of polymeric lubricants has been exploited in bearings that are normally lubricated by a circulating oil system, where there can be a delay in the arrival of the oil at a critical location. The same effect has been used to provide a backup, should the oil supply system fail.

The high occupancy ratio of the void space by the polymer minimize the opportunity for the bearing to "breathe" as temperature change, thereby reducing corrosion caused by internal moisture condensation. Because all ferrous surfaces are very close to the pack, conditions are favorable for using vapor-phase corrosion control additives in the formulations.

Despite these advantages, polymeric lubricants have some specific drawbacks. There tends to be considerable physical contact between the pack and the moving surfaces of the bearing. This leads to increased frictional torque, which produces more heat in the bearing. In conjunction with thermal insulating properties of the polymer and its inherently limited temperature tolerance, the speed capability is reduced. Moreover, the solid polymer is relatively incapable of entrapping wear debris and dirt particles, as compared with grease.

Solid Lubricants

Solid lubricants are used whenever conventional lubricants are not suitable. Extreme environment conditions frequently favor solid lubricants, which can survive temperatures well above the decomposition temperatures of oils. They can also be used in chemically reactive environments. The disadvantages of solid lubricants are their high coefficient of friction, inability to act as a coolant, finite wear life, difficult replenishment, and small damping effect for controlling vibrational instabilities of rolling elements and separator components.

Many common solid lubricants, such as graphite and molybdenum disulfide are layered lattice compounds that shear easily along preferred planes of their structure. Molybdenum disulfide has weak van der Waals forces between sulfur bonds, giving the material a characteristic relatively low coefficient of friction. It also oxidizes at approximately 399 °C (750 °F) in air, and the oxides can be abrasive.

The low friction associated with graphite depends on intercalation with gases, liquids, or other substances. For example, the presence of absorbed water in graphite imparts good lubricating qualities. Thus, pure graphite has deficiencies as a lubricant except when used in an environment containing contaminants such as gases and water vapor. With proper additives, graphite can be effective up to 649 °C (1200 °F). Tungsten disulfide is similar to molybdenum disulfide in that it is a type of layered lattice solid lubricant. It does not need absorbable vapors to develop low shear strength characteristics.

Other "solid" lubricating materials are solid at bulk temperatures of the bearing, but melt from frictional heating at points of local contact, giving rise to a low shear strength film. This melting can be very localized and of very short duration. Soft oxides, such as lead monoxide, are relatively nonabrasive and have a relatively low friction coefficient, especially at high temperatures where their shear strengths are reduced. At these temperatures, deformation occurs by plastic flow rather than by brittle fracture. Melted oxides can form a glaze on the surface, which can either increase or decrease friction, depending on the "viscosity" of the glaze within the contact region. Stable fluorides, such as lithium fluoride, calcium fluoride, and barium fluoride, also lubricate well at high temperatures, but over a broader range than lead oxides.

Metalworking Lubricants

Joseph T. Laemmle, Aluminum Company of American, Alcoa Technical Center

Introduction

DURING METALWORKING OPERATIONS, tool and workpiece surfaces are brought together in the presence of a lubricant. The tool is used to change the shape of the workpiece. Typical examples of tools used in metalworking operations are a drill bit in a drilling operation, a work roll in a rolling operation, and a forging die in a forging operation. In the drilling operation, the workpiece is the object being drilled. In the rolling operation, the workpiece is the strip being reduced, and in the forging operation, the workpiece is the billet being shaped.

Metalworking lubricants have some functions that are common to all metalworking operations, such as

- Controlling friction
- Reducing tool wear
- Improving surface quality of the workpiece

Common functions and special requirements that a metalworking lubricant must meet when used, for example, to roll aluminum that is to be used for food packaging include:

- Control friction
- Reduce roll wear
- Cool roll and strip
- Minimize transfer of aluminum from the strip to the work roll and produce a uniform and stable work roll coating
- Produce an appropriately bright, smudge-free surface
- Produce a surface free of defects caused by lubricant failure
- Protect the sheet surface from water stains and handling marks in downstream operations
- Resist oxidative degradation and bacteriological attack
- Withstand plant environments and resist performance changes due to contamination
- Be compatible with lubricants used in downstream operations
- Prevent corrosion
- Be easily filtered
- Leave residues that do not produce stains upon sitting or annealing
- Ease of maintenance
- No production of objectional odors
- Cost effectiveness
- Easy waste disposal
- Satisfy regulatory requirements, such as those mandated by the Food and Drug Administration (FDA)
- Satisfy hygiene and environmental criteria

The formulation or selection of a modern metalworking lubricant that meets all process and product requirements is often a sophisticated task.

One of the two generally recognized categories of metalworking lubricants is represented by lubricants that are used in operations where metal is removed from the workpiece in order to obtain a desired shape. Examples of metal removal operations include drilling, broaching, turning, grinding, milling, threading, reaming, boring, and sawing. The other category is represented by metal forming lubricants, which are used in operations where metal is plastically deformed to obtain the desired shape of the workpiece. Examples of metal forming operations are hot rolling, cold rolling, foil rolling,

forging, wire drawing, tube drawing, deep drawing, ironing, extrusion, and spinning. Because metalworking lubricants can generally cool, as well as lubricate, they are often called metalworking fluids, metalworking coolants, or lubricant-coolants.

Ferrous metals, such as carbon steel, low-alloy steel, and stainless steel, constitute the majority of fabricated metal products. Large quantities of aluminum are subjected to metalworking operations, with lesser amounts of copper, brass, and titanium also being processed. Other metals subjected to either removal or forming processes include nickel-based alloys, cobalt-based alloys, magnesium, zinc, tin, beryllium, zirconium, tungsten, molybdenum, tantalum, uranium, and vanadium.

Considering the various types of metalworking operations involving various alloys with different sets of requirements, it should be evident that lubricant formulation, selection, and use is a complex process. This article presents an overview of lubricant properties common to metalworking operations, followed by discussions involving lubricant issues that are especially important to metal removal operations and metal forming operations.

Acknowledgements

The author would like to thank both Mr. John Bohaychick, who aided in the preparation of this article, as well as the Aluminum Company of America for permission to publish.

Formulations

Metalworking lubricant formulations commonly consist of a petroleum oil compounded with additives of various types. The oil-additive combination is often emulsified in water. Lubricants that are based on synthetic oils, rather than petroleum oils, are also used. Some metalworking lubricants contain no oil at all, but consist of friction-modifying additives and corrosion inhibitors dispersed or solubilized in water.

Petroleum oils are naturally occurring materials that are refined by processes such as distillation, hydrotreating, solvent extraction, molecular sieving, and dewaxing to give desired properties. Properties include physical attributes, such as viscosity, color, and odor, as well as chemical attributes, such as degree of saturation and freedom from undesirable elements such as sulfur.

Petroleum oils are commonly called mineral oils and are referred to as either bases or base oils in terms of metalworking lubricants. Petroleum oils are generally classified as paraffinic, naphthenic, or aromatic. Paraffinic oils are further classified as either linear paraffins or isoparaffins. The molecular species associated with each classification is illustrated in Fig. 1. Linear paraffins consist of straight-chain hydrocarbons, whereas isoparaffins consist of branched-chain hydrocarbons. Naphthenic oils consist of hydrocarbons that contain five- or six-member ring structures that may be unsaturated, but not aromatic. Aromatic oils consist of hydrocarbons that contain totally unsaturated six-member rings known as either benzene or aromatic rings. Petroleum oils generally consist of a mixture of paraffinic, naphthenic, and aromatic molecules and are classified by the species that predominates. For example, a petroleum oil consisting of 70% naphthenic molecules is called naphthenic. The different classes of oils have different properties. Linear paraffins, for example, have high resistance to oxidative degradation, whereas naphthenic oils are more easily emulsified.

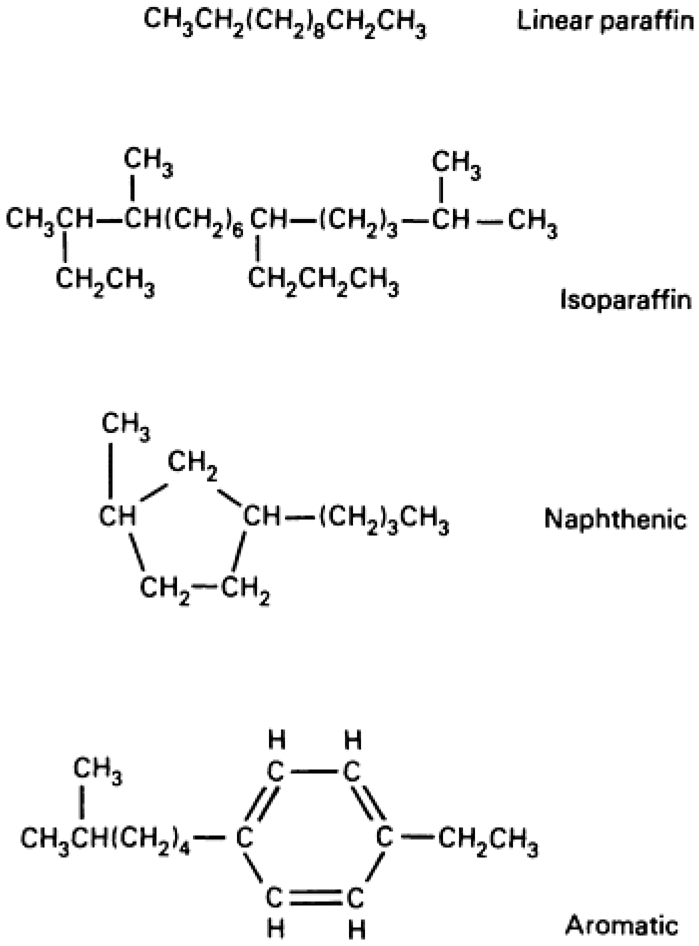


Fig. 1 Examples of molecular species contained in petroleum oils

Synthetic Fluids. Although less commonly used than petroleum oils, synthetic fluids are becoming increasingly popular as bases for metalworking lubricants in cases where their tailored properties can more than make up for their higher cost. Examples of some synthetic fluids are illustrated in Fig. 2. Synthetic hydrocarbons, such as polyisobutylenes and poly-α-olefins, are often referred to as synthetic oils.

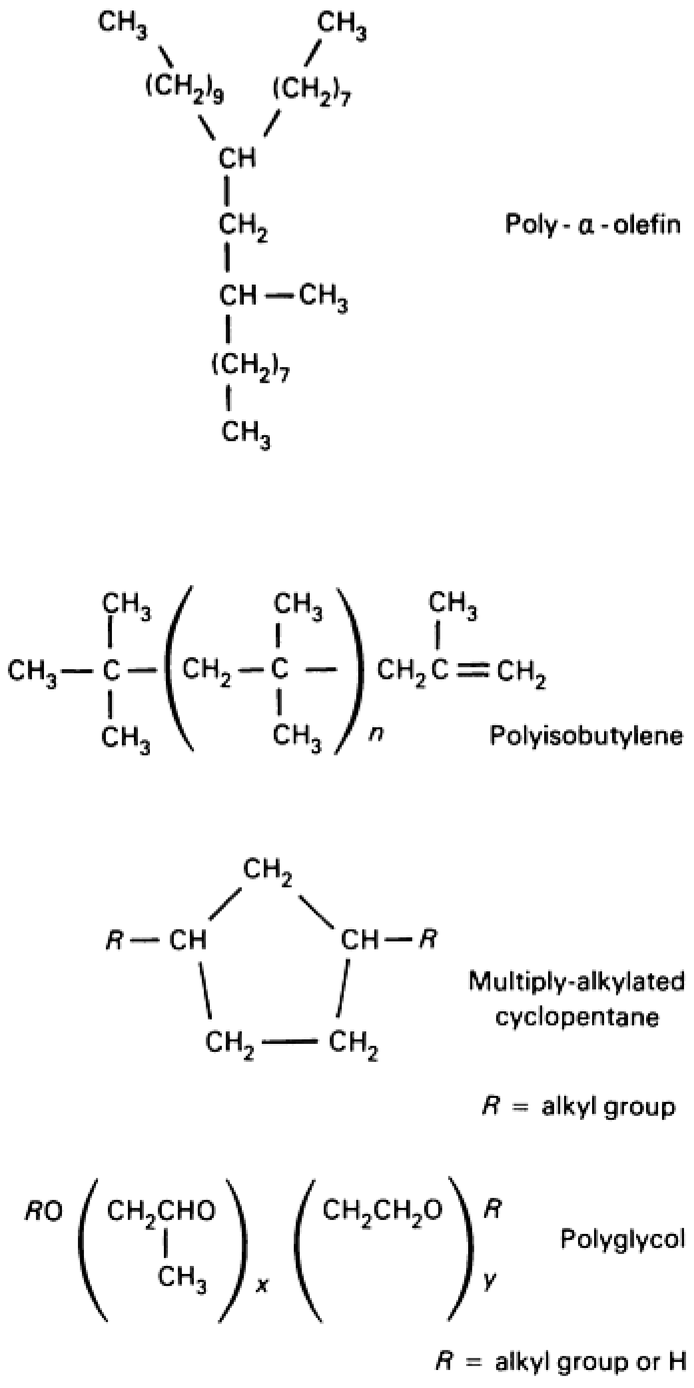


Fig. 2 Examples of molecular structure of four classes of synthetic fluids

Polyglycols are often used in metal removal lubricant formulations. These materials are prepared through polymerization of ethylene oxide and propylene oxide in either random or block fashion. The terminal groups can be either alkyl groups or hydrogen. Properties are controlled by molecular weight, nature of terminal groups, and the ratio of ethylene oxide to propylene oxide. For example, high molecular weight versions have high viscosities. Polyglycols with high ratios of ethylene oxide to propylene oxide tend to be water soluble, whereas those with low ratios tend to be water insoluble. The ratio can be adjusted such that the polyglycol is water soluble at room temperature, but water insoluble at elevated temperatures. This property can be very useful in formulating solutions that are clear in the circulating system, but will separate polyglycol and the additive package from solution at the hot tool-workpiece interface.

Polyisobutylenes can be made in a wide range of viscosities by controlling molecular weight. They have a unique property of depolymerizing at high temperatures. This makes them useful as rolling and drawing oils for ferrous and

nonferrous forming operations where subsequent annealing would produce staining if petroleum oil forming lubricants were used.

Poly-α-olefins, while not as commonly employed in metalworking operations as polyglycols and polyisobutylenes, have properties that make them useful for some operations. They can be used over a wide temperature range, and they produce less hydrocarbon emissions than petroleum oils at similar viscosities. They also are very resistant to oxidative and thermal degradation.

Multiply-alkylated cyclopentanes are a new class of synthetic hydrocarbons that are promising in terms of future formulations. Their properties can be varied over a wide range by varying the number and nature of alkyl groups. They have excellent pour points and viscosity indexes, as well as exceptionally low volatility. A number of synthetic fluid categories beyond those discussed above also exist, but have very limited use as base oils in metalworking operations.

Viscosity

Viscosity is a measure of the resistance of a fluid to flow and is a very important lubricant property. Kinematic viscosity, which represents the resistance of a fluid to flow under gravity, as measured by the ASTM D 445 test, is the preferred method for describing the viscosity of lubricants. In ASTM D 445, a fixed volume of lubricant is allowed to flow through a calibrated orifice that is held at constant temperature. The kinematic viscosity is calculated by multiplying the flow time in seconds by the calibration constant of the viscometer. The correct SI unit of kinematic viscosity is mm^2/s. A centistoke, which is equivalent to 1 mm^2/s, is also commonly used. The standard temperatures for measuring viscosity are 40 °C (104 °F) and 100 °C (212 °F). The viscosity of fluids decreases as temperature increases. Viscosity index is an empirical, dimensionless number that indicates the rate at which lubricants change kinematic viscosity with temperature. The higher the viscosity index of a lubricant, the less rapidly its viscosity changes with temperature.

In metal forming operations, viscosity is a major factor in determining the lubricant film that separates the tool from the workpiece and thus is critical in controlling friction and wear. In metal removal operations, optimum lubricant viscosity must be estimated for a particular operation. Factors to be considered in estimating the optimum viscosity of a metal removal fluid include the capability of the lubricant to penetrate and remain in the contact zone, the durability of the lubricant film, the desired rate of spreading, and the cooling capability.

Additives

In most metalworking operations, the base oil does not totally separate the tool from the workpiece, nor is it desirable to do so. In rolling operations, for example, complete separation of the work roll and strip leads to very low friction and loss of mill control. In metal removal operations, the tool surface must contact the workpiece surface in order for chip formation to occur. In both cases, additives to the base oil are required to modify friction, control tool wear, and protect workpiece surfaces in those areas where tool-workpiece surface contact occurs. Additives that preserve the life of the formulated coolant and that prevent corrosion of the tool, workpiece, and lubricant handling system are also commonly used. In water-based metalworking lubricants, special materials are employed to disperse or solubilize the oil-additive package in water. Additives commonly used in metalworking formulations are described below.

Film-strength additives adsorb on tool-workpiece surfaces and prevent direct metal-to-metal contact and the subsequent welding of asperities and destruction of the workpiece surface. These additives are often called boundary additives, load-bearing additives, oiliness additives, or friction modifiers. Film-strength additives consist of materials with a polar head and a hydrocarbon tail generally containing 10 or more carbon atoms. Compounds that function as film-strength additives include fatty acids, esters, alcohols, amides, amines, and alkyl acid phosphates.

Figure 3, which illustrates how film-strength additives work, shows the results of a crossed-cylinders test in which a stationary aluminum specimen 6.35 mm (0.25 in.) in diameter transversed a circular path on a rotating steel specimen 19 mm (0.8 in.) in diameter at a velocity of 411.5 mm/s (16.2 in./s) and a total load of 8.895 N (2 lbf). Three lubricants were used to lubricate the steel-aluminum contact: two linear paraffins of different viscosities and a linear paraffin of lower viscosity formulated with 5% methyl ester film-strength additive. Higher-viscosity lubricants give thicker lubricant films and, therefore, less tool-workpiece surface contact and lower friction in metalworking operations. This is clearly seen in Fig. 3. Even the higher-viscosity linear paraffin, however, shows much spiking in the frictional force trace, indicating tool-work-piece surface contact with localized asperity welding and associated friction spikes. The addition of the film-strength additive eliminates friction spikes and gives an overall low and smooth frictional force trace.

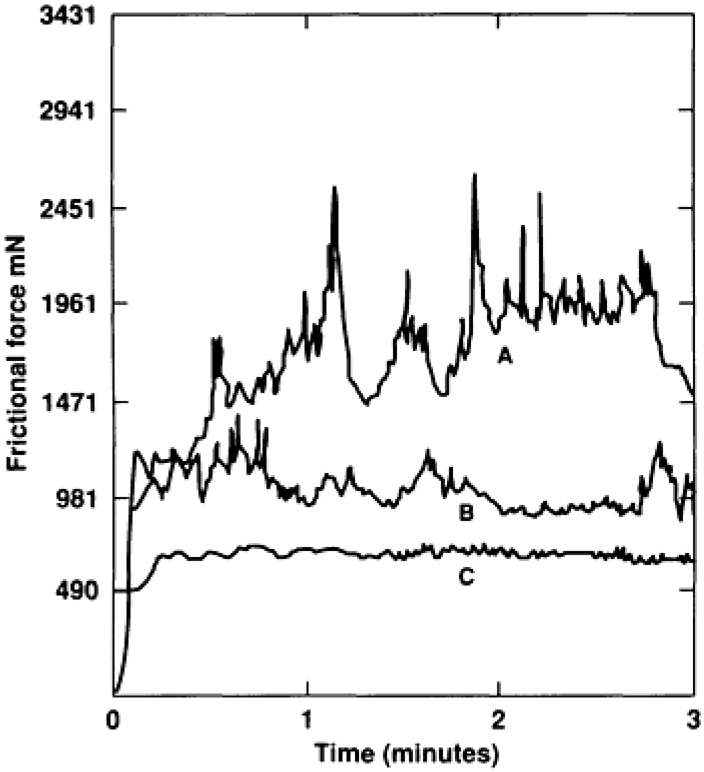


Fig. 3 Crossed-cylinders frictional force showing 1.18 mm^2/s (1.8 × 10^{-3} $in.^2/s$) linear paraffin (curve A), 2.49 mm^2/s (3.9 × 10^{-3} $in.^2/s$) linear paraffin (curve B), and 5% methyl ester in 1.18 mm^2/s (1.8 $in.^2/s$) linear paraffin

Extreme pressure (EP) additives are generally useful in relatively severe metalworking operations involving ferrous metals. The most commonly employed EP additives are compounds containing sulfur, chlorine, phosphorus, or some combination of two or more of these elements. Closely associated with EP additives are so-called "antiwear" additives containing the same elements. Extreme pressure additives function by forming a reaction layer triggered by the high temperatures reached at the tool-work-piece interface in ferrous metalworking operations. These reaction layers reduce friction and wear. Phosphorus-containing compounds produce iron phosphates, iron pyrophosphates, or iron phosphides, depending on the nature of the compound. Sulfur-containing compounds produce iron sulfide, whereas chlorine-containing compounds produce iron chloride. Examples of EP additives include sulfurized triglycerides, chlorinated hydrocarbons, chlorinated esters, phosphate esters, and alkyl acid phosphates. A commonly used class of antiwear additives are the zinc dialkyl dithiophosphates.

A new class of EP additives, called passive extreme pressure (PEP) additives, has been recently introduced. They function by adsorbing a film of carbonate particles at the tool-workpiece interface in metalworking operations. These films have low shear strengths and high melting points. They reduce friction and minimize metal transfer from the workpiece to the tool. Passive extreme pressure additives do not contain phosphorus, sulfur, or chlorine, but are synergistic with sulfur-containing EP additives. They offer advantages over conventional EP additives in that they are less corrosive, are more easily disposed of after use, are low foaming, and are easily cleaned from the workpiece surface. They can be used with both ferrous and nonferrous metals.

Suspended Solids. Solids such as graphite, molybdenum disulfide, metal powders, metal oxides, metal halides, mica, and polytetrafluoroethylene are used as lubricants suspended in either oil or water carriers for certain metal forming operations, such as extrusion and forging. The solids are generally in the form of very finely divided powders held in suspension by either mechanical agitation or emulsifiers, or a combination of both. The oil or water carrier often functions solely to coat the die and work-piece in a uniform fashion, and the solid itself functions as a lubricant. Solids employed as lubricants have the general property of being easily sheared at the tool-workpiece interface. Suspended solids are not generally used in metal removal lubricants, although some types of PEP agents are colloidal suspensions.

Emulsifiers, often called surfactants, are materials that have portions within the same molecule that are hydrophilic, or compatible with water, and lipophilic, or compatible with oil. The hydrophile-lipophile balance (HLB) of an emulsifier is a measure of its tendency to be more compatible with water or oil. The HLB scale runs from 0 to more than 30, with oil compatibility decreasing and water compatibility increasing at higher numbers. Emulsifiers with HLB values higher than 13 form clear solutions in water.

Because emulsifiers have portions of their molecules compatible with water and other portions compatible with oil, they tend to concentrate at the oil-water interface in oil-water mixtures, reduce oil-water interfacial tension, and thereby promote emulsification. The emulsification process is that technique in which oil globules that are larger than colloidal size are dispersed in water. Petroleum oils, animal and vegetable oils, synthetic oils, and waxes each have an HLB value at which they form the most stable emulsion. This is known as the "required HLB" of the material to be emulsified. Required HLB values are different for different materials. For example, the required HLB to emulsify a paraffinic petroleum oil is 10, whereas the required HLB to emulsify castor oil is 14. An emulsifier or emulsifier combination that has the same HLB as that required of the material to be emulsified is chosen if the most stable emulsion is desired.

Emulsifiers are classified as cationic, anionic, amphoteric, and non-ionic. The most common cationic emulsifiers are long carbon chain quaternary ammonium halides. Anionic emulsifiers include alkali metal and amine soaps of long carbon chain fatty acids, and long carbon chain sulfates or sulfonates. Examples of amphoteric emulsifiers include long-chain amino acids and alkyl betaines. Examples of non-ionic emulsifiers include ethoxylated alcohols, ethoxylated fatty acids, and ethoxylated sorbitol. The most commonly employed emulsifiers in metalworking lubricants are the non-ionic and anionic types.

Corrosion inhibitors for ferrous metals are often contained in metalworking lubricant formulations, particularly those that are water based, in order to protect the tool, machine parts, and lubricant system. In some instances, they also protect the workpiece during and after processing. Prior to the 1980s, a commonly used and very effective corrosion inhibitor was sodium nitrite. In the mid-1970s, it was discovered that nitrosamines, which are carcinogenic, were contained in many commercial metalworking formulations. The nitrosamines were found to arise through a reaction between sodium nitrite and secondary amines, particularly diethanolamine. Much work has been one in recent years to find suitable replacements for sodium nitrite.

A number of types of compounds have been developed for the prevention of ferrous metal corrosion in modern metalworking lubricants. These include amine-borates, amine carboxylates, amine alkyl acid phosphates, and sulfonates. None of these materials forms nitrosamines and are therefore more acceptable.

Corrosion inhibitors for nonferrous metals are important in some metalworking operations. Benzotriazole is an effective corrosion inhibitor for copper or brass. Cobalt corrosion inhibitors, such as tolyltriazole, are used to prevent cobalt leaching in those operations employing cobalt-cemented tungsten carbide tools.

Oxidation inhibitors, also called antioxidants, are formulated into metalworking lubricants to minimize oxidative degradation of the lubricant into acidic products that tend to form sludge and corrode metal surfaces. Oxidative degradation involves molecules containing unpaired electrons called free radicals. These radicals are transformed into degradation products while transforming other lubricant molecules into more free radicals in a chain reaction. Small concentrations of oxidation inhibitors functioning as free radical scavengers intercept the unpaired electrons and break the chain. Oxidation inhibitors are not normally used at levels above 0.5%. Commonly used oxidation inhibitors include hindered phenols, such as butylated hydroxytoluene and butylated hydroxy anisole, and secondary aromatic amines, such as phenyl naphthyl amine. Because the rate of oxidation occurs much more rapidly at elevated temperatures, oxidation inhibitors are especially useful for hot metalworking operations, such as the hot rolling of either ferrous or nonferrous metals.

Defoamers. Lubricant foaming can have a deleterious effect on metalworking operations. Foam inhibits tool-workpiece interface cooling, creates a mess if it overflows sumps and tanks, and can even lead to lubricant starvation if excessive amounts of the lubricant mass are in the form of foam. Nearly all water-based metal-working lubricants contain emulsifiers, which not only lower oil-water interfacial tension, but also lower the surface tension of the lubricant compared to water. The surfactants concentrate at the air-fluid interface and form an elastic film that expands, but does not rupture, as air is introduced. Foam is created when air is injected into the lubricant either through spraying at the metalworking operation or through circulation of the lubricant through the lubricant handling system. Small additions of chemical agents, called anti-foams or defoamers, can drastically reduce or eliminate foams. A good defoamer must have the right combination of dispersibility and surface tension. It should spread throughout the system without dissolving in it, and it should spread over the foam surfaces. When this happens, the defoamer acts at the gas-fluid interface to collapse

the elastic film of the fluid, thereby allowing it to release the air and drain. Silicones are very effective defoamers. Their major drawback is that if the workpiece is to be coated or painted in a subsequent operation, adhesion may be adversely affected. Nonsilicone defoamers include long-chain alcohols, certain triglycerides, and water-insoluble polyglycols.

Antimicrobial agents are materials designed to inhibit the growth of bacteria, fungi, and yeast in metalworking lubricants. All water-based metalworking lubricants are vulnerable to attack by one or more of these agents; even oil-based lubricants containing small amounts of water as a contaminant can be degraded by microbes. Attack of metalworking lubricants by bacteria leads to one or more of the following: buildup of acidic materials, corrosion of machinery and tools, destruction of additives, objectionable odors, and loss of stability in emulsions. Growth of fungi can lead to slimy material coating the machinery and tools, as well as the clogging of pumps and filters. Bacteria, fungi, and yeast are often monitored on a regular basis through commercially available simple culture techniques, and when counts reach a certain level, there is cause for alarm.

Microbes can generally be controlled at acceptable levels through use of antimicrobial agents known as biocides and fungicides. Standard practice often calls for the addition of two different biocides to the metalworking lubricant at regular intervals in an alternating fashion, in order to guard against microbes developing an immunity to one of them, resulting in an uncontrolled infestation.

Although many types of biocides exist, two of the most common are phenolic materials and formaldehyde-release agents. Phenolic materials, such as 2,4,5-trichlorophenol, destroy bacteria directly. Materials such as 1,3-di(hydroxy-methyl)-5,5-dimethyl-2,4-dioxoimidazole, upon being added to water-based metalworking lubricants, release formaldehyde slowly to keep bacteria in check. Materials such as 2,2-dibromo-3-nitrilopropionamide are useful for controlling bacteria, fungi, and yeast.

There are over 50 commercially available antimicrobial agents. In choosing the proper one, parameters such as the required concentration, effect on emulsion stability, and regulations concerning discharge into waste streams must be taken into account. Finally, antimicrobial agents are designed to destroy living organisms. They all display some degree of toxicity toward humans and should be handled with caution.

Metalworking Lubricant Types

Four commonly used types of metalworking lubricants are illustrated in Fig. 4. Straight metalworking oils, often simply called straight oils, are given this term because they are not mixed with water prior to use. Emulsions are mixtures of either simple or compounded oils with water, stabilized by the use of emulsifiers. Emulsion droplets are similar in size to, or larger than, the wavelength of visible light; hence, emulsions appear milky white. Microemulsions can be similar to emulsions in composition, but, through emulsifier choice, have oil droplet diameters that are much smaller than the wavelength of visible light and therefore appear transparent. Micellar solutions are similar to microemulsions, except that they contain no oil.

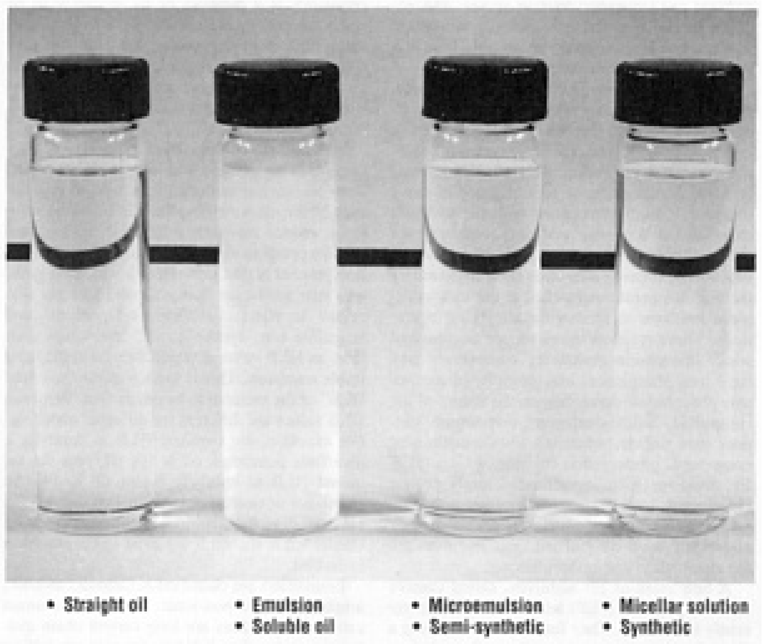


Fig. 4 Commonly used metalworking lubricants

Petroleum oils, as previously defined, are naturally occurring materials that are refined through processes that separate the crude substances into various molecular fractions and remove impurities. In some cases, they are hydrogenated or subjected to reforming catalysts, but are not otherwise chemically altered. The chemical reaction of smaller molecules produces the larger molecules of synthetic fluids, such as poly-α-olefins and polyisobutylenes. In metalworking lubricants, the term synthetic is often used to describe either transparent micellar solutions (Fig. 4) or true solutions containing no petroleum oils. Unfortunately, this dual definition for synthetic lubricant has led to confusion. It should be obvious that if the definition of synthetic lubricant is one that contains a preponderance of man-made materials, then the straight oil, emulsion, or microemulsion of Fig. 4 is, strictly speaking, synthetic, if the base oil is a synthetic fluid such as poly-α-olefin, rather than a petroleum oil.

Straight oils are generally petroleum oil fractions that are normally formulated with either film-strength additives or EP additives or a combination of both. They generally provide excellent friction reduction and workpiece surface finish, good corrosion protection, and a long service life. Straight oils containing certain EP additives will stain nonferrous metals, such as copper, and are commonly referred to as staining oils. The major disadvantage of straight oils is their poor capability for heat removal, compared to water. In addition, straight oils with low flash points, coupled with the high temperatures often encountered in metalworking operations, can create fire hazards. Straight oils are commonly used in metalworking operations where lubrication is a major factor and cooling is a minor factor. Examples of such operations include low to moderate speed metal removal operations where accuracy, tolerance, and workpiece finish are important, and metal forming operations such as aluminum foil rolling where strip surface quality is highly important.

Emulsions. In commonly used metalworking emulsions, oil globules are finely dispersed in water, and this oil-water combination is employed as a lubricant-coolant. In these types of emulsions, oil is said to be the dispersed phase, and water, the continuous phase. Oil-in-water mixtures are thermodynamically unstable; that is, their state of lowest free energy is total separation. Because of this, oil tends to separate, and emulsifiers are added to stabilize the emulsion. Emulsifiers concentrate at the oil-water interface and inhibit coalescence of oil globules. This is illustrated in Fig. 5. The structural formula of sodium oleate, an anionic emulsifier, is shown in Fig. 5(a). A simplified "straight pin" is depicted in Fig. 5(b). An oil-in-water emulsion stabilized by sodium oleate is shown in Fig. 5(c). The hydrocarbon chain of sodium oleate is compatible with the oil globules and penetrates them. The carboxylate head of sodium oleate is compatible with water and lies at the surface of the oil droplet penetrating into the water phase. Because the carboxylate head carries a negative charge, the surface of each oil droplet is negatively charged, and since like charges repel, the oil droplets tend to stay dispersed. Because the state of lowest free energy of the emulsion is still total separation, the emulsion is said to be kinetically stabilized.

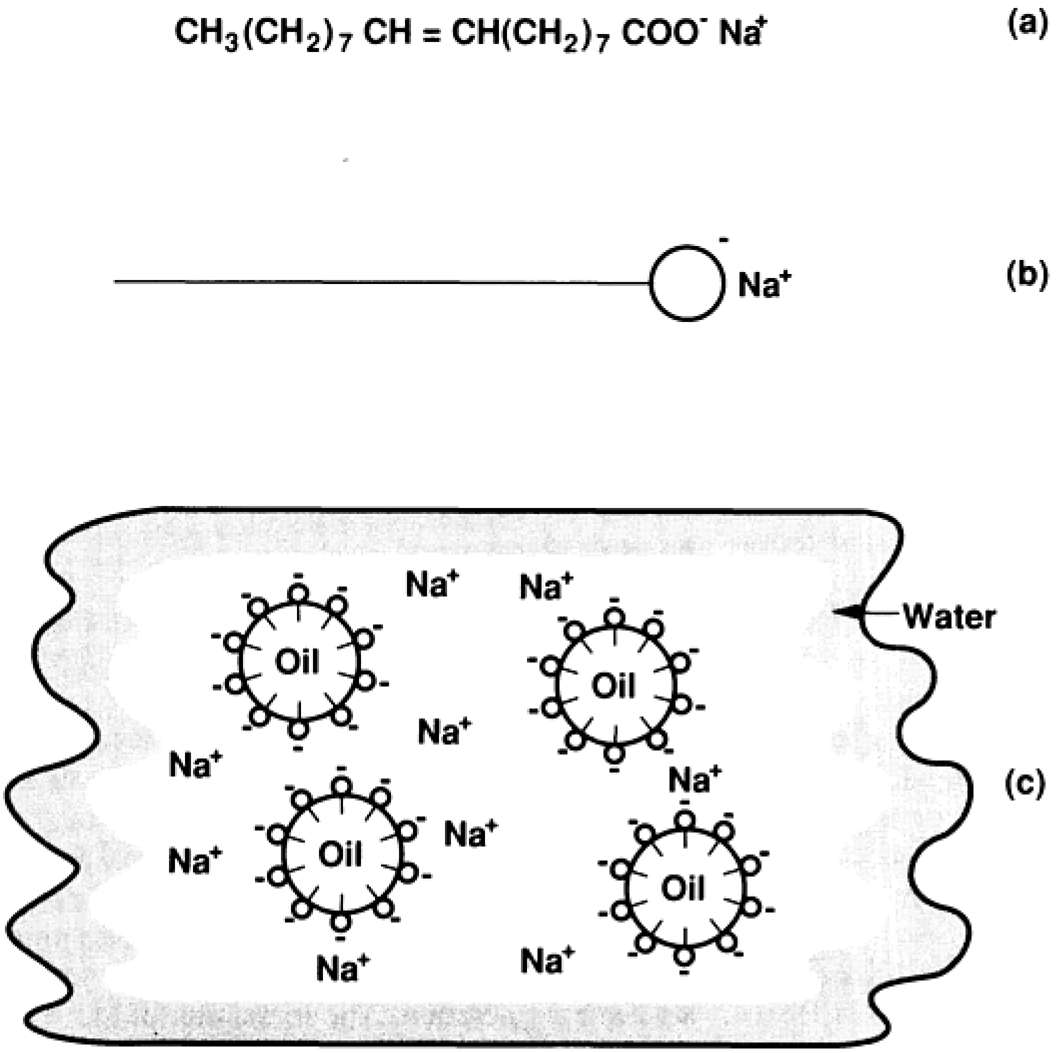


Fig. 5 (a) Sodium oleate. (b) "Straight pin" depiction of sodium oleate. (c) Oil-in-water emulsion stabilized by sodium oleate emulsifier

Emulsions can vary in stability over a wide range, depending on the nature of the oil phase and the nature and concentration of the emulsifier package. Emulsion oil globule size can vary between about 0.2 μm (8 μin.) to as high as 10 μm (400 μin.) or more. Therefore, emulsions appear as off-white to white opaque solutions. The globule sizes within a given emulsion are polydispersed; that is, they vary over some distribution. Stable emulsions have smaller average globule size distributions than unstable ones.

A major factor in lubricating with emulsions is the availability of the oil phase to lubricate. Two factors control oil availability: the emulsion stability and the concentration of oil in the emulsion, which is often called "percent oil." In general, the less stable the emulsion and the higher the percent oil, the greater is the availability of oil for lubrication. Unfortunately, the less stable an emulsion is, the higher is the tendency for stability to change, sometimes rapidly, over time. This can lead to undesirable instability in some metalworking operations, such as rolling. Also, as the percent oil in an emulsion increases, cooling capability decreases. Therefore, the stability and percent oil in an emulsion must be carefully balanced to satisfy the lubrication and heat removal needs of a particular metalworking operation.

The fact that emulsions are kinetically, rather than thermodynamically, stable leads to other factors in their behavior. One of these is called the emulsion "batch life." New emulsions are generally the most stable and have the least oil available for lubrication. Metalworking operations are often not optimal when a new emulsion batch is introduced. Over time, as debris is generated in the emulsion, providing nucleation sites for oil globule coalescence, and as emulsifiers are depleted, the emulsion becomes less stable and performs at its best. At yet a later time, the emulsion becomes so degraded and unstable as to be rendered useless and is discarded. A new batch is introduced, and the process repeats itself. In general,

emulsions that are initially less stable have shorter batch lives. A second factor is the care that must be taken upon introducing foreign substances into the emulsion. For example, introduction of a biocide to combat a microbial infestation or contamination by acids or bases can greatly affect emulsion stability and therefore the consistency of the metalworking operation.

Metalworking lubricant emulsions are often complex mixtures of emulsifiers, film-strength additives, oxidation inhibitors, corrosion inhibitors, and coupling agents. Coupling agents are generally lower molecular weight diols and triols that aid in the initial emulsification. Emulsions can also contain various mixtures of EP additives. The formulated oil mixture is called the concentrate and is added to water with agitation to form the emulsion. The quality of the water is extremely important. Distilled or deionized water should be used whenever possible. Metalworking emulsions generally operate at levels between 5 and 10% oil.

Metalworking emulsions are maintained in a variety of ways. The percent oil is determined by breaking the emulsion in a graduated bottle with an acid or salt solution, and it is maintained by adding new concentrate during the life of the batch. Nonemulsifiable tramp oils (those that have leaked into the metalworking lubricant) are skimmed off, and the emulsion is subjected to continuous filtration to remove fine debris. In the case of metal removal fluids, chips are often removed by mechanical means. Microbe levels are monitored and controlled by the addition of appropriate antimicrobial agents at prescribed intervals. Emulsions have good lubricating and heat removal qualities and are widely used in most metal removal operations and many metal forming operations.

Microemulsions are clear-to-translucent solutions containing water; a hydrophobic liquid, that is, an oil phase; and one or more emulsifiers, which are often referred to as surfactants and co-surfactants. Microemulsions in which water is the continuous phase and oil is the dispersed phase are called oil-in-water microemulsions and are the type generally used as metal-working lubricants. Microemulsions employed as lubricants are commonly called semisynthetic fluids.

Several parameters differentiate microemulsions from emulsions. Most importantly, microemulsions are thermodynamically stable; that is, the state of lowest free energy is dispersed rather than separated. Therefore, the stability problems associated with emulsions are nonexistent. Microemulsions remain stable indefinitely, as long as they are maintained in appropriate ranges of pH, oil-to-water ratio, and temperature. These ranges may be very narrow to very broad, depending of the nature of the microemulsion. The diameters of the dispersed oil globules in microemulsions range from about 0.01 to about 0.2 μm (0.4 to 8 μin.), depending on the nature of the oil and the types and concentrations of emulsifiers. This small oil globule size is the feature that makes them appear clear to translucent. Additionally, the oil globule diameters are much more uniform in microemulsions than they are in emulsions.

Microemulsions are generally produced in the metalworking environment by adding a concentrate to water with agitation. The concentrate generally contains oil, the emulsifier package, film-strength additives, corrosion inhibitors, biocides, and in some cases, EP additives. Dilutions range from about a 10:1 ratio of water to concentrate to as high as 60:1. Lower dilutions are used in operations were lubrication is more important, whereas higher dilutions are used where cooling is more important. Concentration is commonly determined with a hand-held refractometer and a calibration chart that relates the instrument reading to concentration.

Microemulsions offer good resistance to corrosion and to microbial attack, as well as excellent stability and cooling. They suffer from higher initial cost, difficulty of disposal, and a stronger tendency to foam. Microemulsions formulated with fatty acid soap-type emulsifiers tend to degrade rapidly in hard water, because of the formation of insoluble calcium and magnesium carboxylates.

Micellar Solutions. When emulsifier molecules are dissolved in water, they tend to aggregate into larger units called micelles. Micelles are spontaneously formed because of the fact that the lipophilic portion of the emulsifier molecule tends to aggregate in the interior of the micelle, whereas the hydrophilic portion tends to penetrate into the water phase. Figure 6 illustrates, in two dimensions, the relation between a molecule of a typical anionic emulsifier, sodium dodecyl sulfate, and the spherical micelle that it forms in water. Micellar solutions used as metal-working lubricants contain neither petroleum oils nor synthetic hydrocarbons. They contain film-strength additives, EP additives as appropriate, and corrosion inhibitors solubilized within the interior of the micelles. Because micelles have diameters typically between about 0.005 and 0.015 μm (0.2 and 0.6 μin.), micellar solutions are transparent to the eye and, like microemulsions, are thermodynamically stable. Because virtually all of the components of micellar solutions are obtained by chemical synthesis, they are often referred to as either synthetic lubricants or chemical coolants.

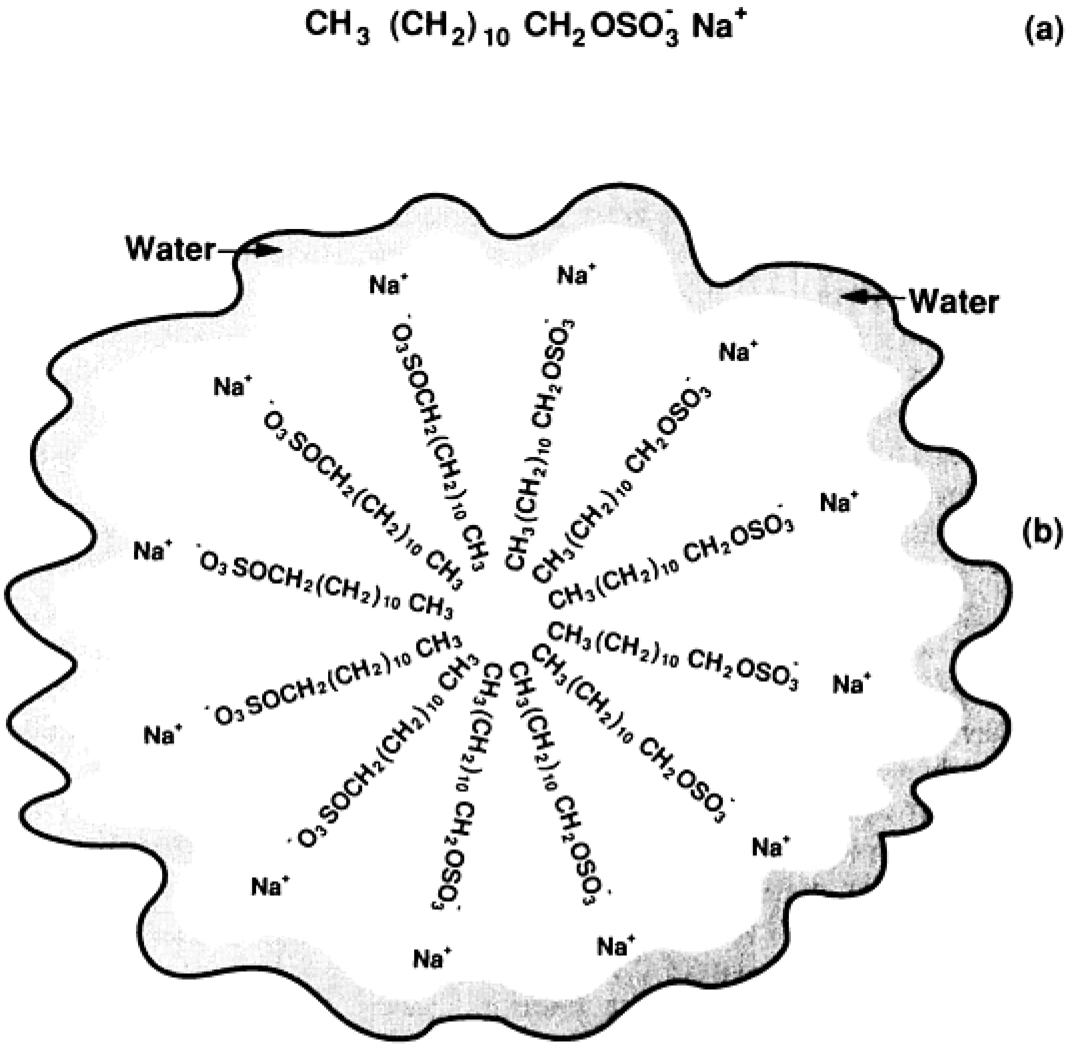


Fig. 6 (a) Molecule of sodium dodecyl sulfate. (b) Sodium dodecyl sulfate micelle in water

Like emulsions and microemulsions, micellar solutions for metalworking are generally formed by the addition of a concentrate to water with agitation. Dilutions typically range from a water-to-concentrate ratio of 10:1 to about 50:1, depending on the application. As in the case of microemulsions, concentration is determined by refractive index.

Micellar solutions that contain no alkali metal soaps or amine fatty acid soaps show good stability in hard water. They are more resistant to microbial attack than either emulsions or microemulsions. They can be formulated to reject tramp oils, which can then be skimmed and collected for disposal or recycling. They have excellent cooling capability, provide excellent corrosion control, and have a long useful life.

On the downside, because micellar solutions tend to cost more initially, a total cost-benefit analysis should be performed. Additionally, because they are highly fortified with emulsifiers, foam can be a real problem. Antifoaming agents can be added to control foam. Micellar solutions, in general, have lower lubricating capability than other types of metalworking lubricants. This limits their applications to those metal removal operations with low tool pressures and high tool speeds where cooling is of paramount importance. In such operations, tool life can be extended as much as 250 ° by using micellar solutions, compared to straight oils. One final drawback is waste disposal. It is often very difficult to separate the organic materials from water, because of their nature. The organic materials tend to remain soluble over wide ranges of pH, temperature, and salt concentration. It is often necessary to resort to sophisticated techniques, such as reverse osmosis, for disposal.

True solutions differ from micellar solutions in that the molecules of active substances do not form micelles when dissolved in water. Rather, each ion is solvated by water molecules. Because essentially all film-strength additives and EP

additives either form micelles or require micelles to dissolve them, true solutions offer lubricating qualities that are little better than those of water.

True solutions are used in cases where cooling is the only consideration. In these cases, their advantages are low cost, very high cooling, stability, low foam, and a very long life. Typical true solutions often contain nothing more than a corrosion inhibitor, such as sodium nitrite, in water.

Solid-Lubricant Suspensions. Lubricants for specialized uses often contain solids in the form of finely divided powders suspended in a liquid carrier, such as oil or water. The liquid carriers may also contain soluble additives of the classes previously mentioned. One of the most common suspended substances is colloidal graphite, with specific surface areas that often exceed 100 m^2/g (3×10^4 ft^2/oz). It is used extensively in hot forging and extrusion of both ferrous and non-ferrous metals. Molybdenum disulfide is another commonly used suspended solid. Both graphite and molybdenum disulfide are compounds that possess layered crystal structures with weak forces bonding the layers together so that they are easily sheared. They function by plating onto tools and workpieces such that the weak shear direction is parallel to the surfaces. As the tool and workpiece surfaces are brought together, they form a solid film, preventing tool-work-piece contact and shear along the weak shear plane, thereby reducing friction.

Other types of solids that are suspended as fine powders in some metalworking lubricants include mica, polymers such as Teflon, certain metal oxides, and glasses. Mica has a layered structure and functions in a way similar to graphite. Polymers mechanically separate metal surfaces, lower friction, and reduce metal transfer. Hard metal oxides, such as aluminum oxide, have good wear resistance but high friction coefficients. Soft oxides, such as lead II oxide, give relatively low friction coefficients that decrease at higher temperatures, where the mechanism of deformation changes from fracture to plastic flow. Glasses are suspended in lubricants for use in metalworking operations at high temperatures, where they soften on hot die and work-piece surfaces and function as parting agents of low shear strength.

Solids can be kept in suspension by using surface active agents, mechanical agitation, or both. Problems can arise if not enough care is taken and the solid is allowed to "settle out." Also, metalworking lubricants containing suspended solids tend to produce buildup on tools and workpieces that are difficult to clean. These problems can be minimized by the appropriate formulation.

Metal Removal Lubricants

A metal removal operation is shown in Fig. 7. In some cases, such as turning, the workpiece is moved against a stationary tool. In other cases, such as drilling, the tool is moved against a stationary workpiece. Either operation results in essentially the same type of metal removal mechanism. The tool cuts into the workpiece, resulting in the formation of a chip. Workpiece metal is deformed in the metal deformation zone, resulting in about 65% of the heat generated in the operation. The remainder of the heat is generated by friction between the tool and the chip, and the tool and the workpiece. The lubricant penetrates the shear zone and reduces heat by reducing friction and carrying heat away from the tool and workpiece. Heat can also be reduced by increasing the shear angle, thereby reducing the amount of metal deformation that occurs.

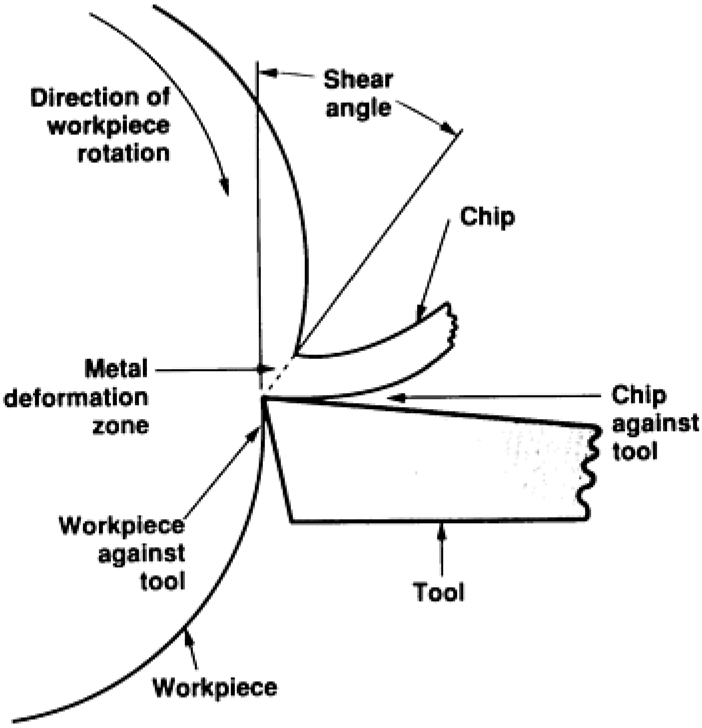


Fig. 7 Metal removal process

One problem that occurs in metal removal, especially with ductile metals, is known as "built-up edge." This problem results when some of the workpiece welds to that part of the tool that is in contact with the chip. Portions of the built-up edge eventually detach, come between the tool and the workpiece, and blemish the workpiece finish. Built-up edge is controlled through film strength and EP additives, which react at the interface to prevent welding, and by choosing an appropriate cutting speed.

In addition to reducing friction and removing heat, the metal removal fluid must flush away chips and debris from the metalworking interface. Unremoved chips can retard the progress of the tool and damage the workpiece. Because a significant amount of the heat generated is located in the chip, removal of the chip is a major factor in reducing heat buildup. An important function in prolonging tool life is reducing and removing heat from the operation.

Selection of metal removal lubricants depends on the operation and on the type of metal composing the workpiece. In operations where low speeds and relatively deep cuts are used, or where workpiece finish and tolerances are important, straight oils are often used. In these cases, lubrication plays a more important role than cooling. Conversely, in high-speed operations with relatively shallow cuts, cooling is most important, and microemulsions or micellar solutions are often employed. Many producers of commercial metal removal fluids provide a fluid selection table with their product literature, such as that shown in Table 1. The table is a matrix with metal removal operations listed in order of decreasing severity on one axis, and various workpiece materials listed on the other axis. The matrix is then filled in with the lubricants recommended for a particular operation and workpiece material. In those cases where the lubricant is dispersed or solubilized in water prior to use, the table often gives the recommended oil:water dilution ratio. In general, straight oils would be more commonly recommended for operations at the top of the table, microemulsions and micellar solutions more commonly recommended for operations toward the bottom, and emulsions recommended over a wide range of operations. However, the recommendations are general, and instances can be found where almost every type of lubricant has been employed satisfactorily in almost every type of operation.

Table 1 Organization of typical metal removal fluid selection chart

Operation severity	Cutting speed	Operation	Ferrous metals	Nonferrous metals
HIGH		Broaching: internal		
		Broaching: surface		
		Threading: pipe		
		Tapping: plain		
		Threading: plain		
		Gear shaving		
		Reaming: plain	Recommended lubricants	
		Gear shaping	and	
		Drilling: deep	recommended dilutions	
		Milling: plain		
		Milling: multiple cutter, hobbing		
		Boring: multiple head		
		Drilling		
		Turning: single-point tools, form tools		
		Sawing: circular, hack		
		Grinding: thread and form		
	HIGH	Grinding: plain		

In addition to this type of table, commercial suppliers furnish information about each lubricant in specification sheets. These sheets customarily contain: recommended uses; physical properties, such as viscosity and specific gravity; staining tendencies on nonferrous metals; and chemical data, such as percent sulfur, chlorine, and fat. Information on materials that are not contained is also provided, such as "contains no nitrites, phosphorus, chromates, or heavy-metal salts." For those materials that are diluted with water prior to use, waste disposal information, such as chemical oxygen demand, biological oxygen demand, oil and grease, and alkalinity, is often given for a particular ratio of water to oil. In addition, commercial suppliers provide material safety data sheets for each lubricant in their product line. These describe potential hazards associated with the use of the lubricant and safe handling procedures, such as required protective clothing like safety glasses and gloves.

Lubricant Application. In most metal removal operations, the lubricant floods both the tool and the workpiece. The lubricant is supplied through high-volume low-pressure spray nozzles to maximize cooling and minimize splashing and foam. The lubricant is customarily directed into the contact zone between the tool and the workpiece. The lubricant is also directed at other positions on the workpiece, where appropriate, to enhance cooling. In Fig. 7, for example, lubricant would be directed into the contact zone to reduce friction, prevent metal transfer, and facilitate cooling. Lubricant would also be sprayed over the back of the chip to further enhance cooling.

Lubricant Maintenance. Lubricants can be maintained through a variety of practices. In water-based coolants, tests for percent oil, pH, tramp oil contamination, suspended solids, microbes, and corrosion are made on a regular basis, and corrective action is taken as required. For example, an increase in pH beyond a specified range might signal contamination by a very basic substance. On the other hand, a decrease in pH might be due to oxidation, microbial infestation, or contamination by an acidic substance. Corrective action could involve adding an oxidation inhibitor, adding a biocide, or finding and eliminating sources of contamination.

During their use, metal removal fluids become contaminated with metal chips and fine debris, such as metal particles and insoluble metal oxides. If not removed, these contaminants can accelerate degradation of the coolant by promoting oxidation and, in the case of emulsions, providing nucleation sites to destabilize the coolant. These contaminants can also damage the workpiece when circulated back to the contact zone. Metal chips are generally allowed to settle and are then removed. The use of centrifuges can accelerate this process. Fine debris is usually removed by filtration, often in two stages. The first stage involves filtration by paper or cloth to remove coarser debris, followed by filtration through filter aids, such as diatomaceous earth or fine volcanic ash, to remove fine debris.

The leakage of hydraulic oils and other lubricating oils (tramp oils) into a metal removal fluid is very detrimental. In the case of straight oils, tramp oil contamination can change the viscosity and dilute the additive package. In the case of emulsions, tramp oils are often emulsified and increase the oil globule diameters and destabilize the emulsion. Some emulsions and many microemulsions and micellar solutions reject tramp oils. In these cases, the tramp oils are removed by skimming.

Reclamation and Disposal. In metal removal operations, lubricant adheres to the chips and is available for reclamation. The chips are moved to a central location with care to segregate chips that come from operations employing different lubricants. The oil can be collected simply by allowing the lubricant to drain from the chips and collecting it. The use of a centrifuge accelerates the process and results in more complete removal of oil. The oil is then cleaned by filtration, as required, and returned to operation.

Most metal removal fluids in use will reach a point in time at which they can no longer be maintained and must be disposed of. No metal removal fluids should be released into the environment without prior treatment. Petroleum oils, natural fats, and greases, as well as synthetic organic materials, are contained in most metal removal fluids. These materials, when released into streams and rivers, float on the water and thereby slow the adsorption of oxygen from the air into the water. In addition, they consume oxygen in the water through direct oxidation and by promoting the growth of bacteria that consume oxygen as they metabolize. Because aquatic organisms require oxygen for survival, they will die if any substantial amount of that oxygen is depleted by the processes noted. There are strict state and federal laws that regulate disposal of metal removal fluids into streams and rivers, with heavy fines for violators.

Spent straight oils are often added to heavier fuel oils and burned to generate heat or power. Emulsions and some microemulsions, as a first step, are treated with acids or salts to separate the organic phase, which is skimmed off. In a second step, the water is then sent to aerated tanks containing aerobic bacteria, where remaining organic materials are consumed. Most plants using large amounts of metal removal fluids have facilities to carry out at least the first step. The water from first-step treatment is often clean enough to be sent to municipal waste treatment systems, where the second step occurs. Small amounts of microemulsions and micellar solutions that resist breaking in the first step can generally be sent directly into the second step as long as their organic content does not overwhelm the system. A third step is sometimes used when the effluent from the second step does not meet water quality standards for discharge. This third step may include such processes as reverse osmosis, chemical oxidation, or oxidation by ozone. Waste from more concentrated microemulsions and micellar solutions can be effectively treated by these methods.

A complete metal removal fluid program, including fluid selection, maintenance, handling, reclamation, and waste disposal, is a vital part of a metal removal operation. Such a program can increase profitability by reducing lubricant costs, allowing increased feeds and speeds, reducing tool wear, and improving workpiece finish, while being environmentally responsible.

Metal Forming Lubricants

In metal forming operations, the desired shape of the workpiece is obtained through plastic deformation. Most metal forming operations employ liquid lubricants consisting of petroleum oil or synthetic oil fortified with additives. These lubricants form films that partially or completely separate the tool from the workpiece, thereby reducing friction and minimizing metal transfer. In many metal forming operations, cooling is also desirable. Therefore, the use of emulsions, and in certain cases, microemulsions and micellar solutions, is common. Certain metal forming operations also use solid suspensions.

Lubricant films in metal forming operations are either "wedge films" or "squeeze films." Wedge films result when two nonparallel surfaces converge under relative motion in the presence of a lubricant. Both surfaces may be in motion, as in the case of rolling, or just one surface may be in motion, as in the case of wire drawing. Squeeze films result when two parallel surfaces approach each other with a liquid lubricant between them. An example of a metal forming operation that develops a squeeze film is an open die forging method called "upsetting."

Wedge Films. Most metal forming operations that employ liquid lubricants develop wedge films. When surfaces converge under relative motion, the lubricant is swept along the moving surface. As the surfaces approach each other, a gap in the form of a wedge develops. Consequently, the entry of the gap is greater than the exit. As lubricant molecules approach the gap entry, they are slowed because the volume they are permitted to occupy is decreasing. Because the number of lubricant molecules that enter and exit the gap must be the same, the molecules exiting the gap are moving faster than the molecules entering the gap, and indeed are moving faster than those molecules prior to approaching the gap entry. The forces that cause the entry lubricant molecules to decrease in velocity and the exit lubricant molecules to increase in velocity also act to push the surfaces apart.

In metal forming operations, the minimum separation between the tool and workpiece surfaces is referred to as the lubricant film thickness, h. For a given operation, h increases as the velocity of the operation and the viscosity of the lubricant increase, and decreases as the force driving the converging surfaces together increases. In rolling, for example, h

increases with the entry velocity of the strip and with the viscosity of the rolling lubricant, but decreases at a given viscosity and entry velocity as reduction increases. Because the viscosities of lubricants increase dramatically at high pressures and decrease at high temperatures, the viscosity that in part controls *h* is that viscosity in the tool-work-piece inlet zone. The viscosity of naphthenic oils, for example, increases at a greater rate with pressure than does that of linear paraffins. Therefore, in a metal forming operation, a naphthenic oil will produce a larger *h* than a linear paraffin with an identical viscosity at 40 °C (104 °F) and an identical viscosity index.

Squeeze films are formed when two parallel surfaces are brought together at some velocity. In operations such as upsetting, squeeze films increase in thickness as both lubricant viscosity and approach velocity increase. Higher viscosity retards lubricant flow from the contact zone, and high approach velocities seal the tool-workpiece edges more effectively, thereby entrapping the lubricant between the tool and workpiece. As workpiece deformation occurs and the surface expands, the squeeze film becomes thinner. Therefore, the thickness of squeeze films is proportional to viscosity, velocity, and force in the same way as that of wedge films, but for different reasons.

Lubrication regimes are defined by which attribute of the lubricant supports the load, that is, modifies friction and reduces metal transfer. In full fluid film lubrication, there is no contact between the tool and workpiece, and force is transmitted from the tool to the workpiece through the lubricant film. In this case, the lubricant film supports the load. In the thin-film regime, there is partial contact between the tool and the workpiece, and both the lubricant film and the film strength or EP additives support the load. In the boundary regime, contact between the tool and workpiece is essentially the same as if there were no lubricant film at all, and the film strength or extreme pressure additives alone support the load.

Rules of thumb exist that relate the ratio, λ, of film thickness to the combined roughnesses of the surfaces, σ, for rigid surfaces. The combined roughness is defined as the sum of the random roughness amplitudes of the two surfaces measured from their average levels. In general, for $\lambda = h/\sigma \leq 0.5$, operation is in the boundary regime. For $0.5 < \lambda \leq 3$, operation is in the thin-film regime, and for $\lambda > 3$, operation is in the full fluid film regime. In metal forming operations where the workpiece is being plastically deformed and the tool is often elastically deformed at the contact zone, λ for each regime is smaller than the values given above. Friction is highest when operating in the boundary regime and lowest in the full fluid film regime. Many metal forming operations start up in the boundary regime and, as operation velocity and therefore thickness increases, move into the mixed film regime and even into the full fluid film regime. Many operations, such as rolling, operate in the thin-film regime when at operational run speed, and a proper balance of lubricant viscosity and additive package is essential.

Friction. Control of friction is very important in most metal forming operations. Figure 8(a) depicts an upsetting operation where a cylindrical billet (B) is being deformed between two dies (D) with an applied force (F). The arrows on the die denote the direction of applied force, whereas the arrows on the workpiece denote the direction of frictional force. Figure 8(b) shows what the deformed billet would look like if friction were absent. The maximum deformation is obtained for a given force, and the walls of the deformed billet are straight. Figure 8(c) shows what the deformed billet would look like in the presence of friction. Because frictional force opposes the flow of the workpiece, at the same applied die force, less deformation occurs. Additionally, the sides of the deformed billet are bulged, rather than straight.

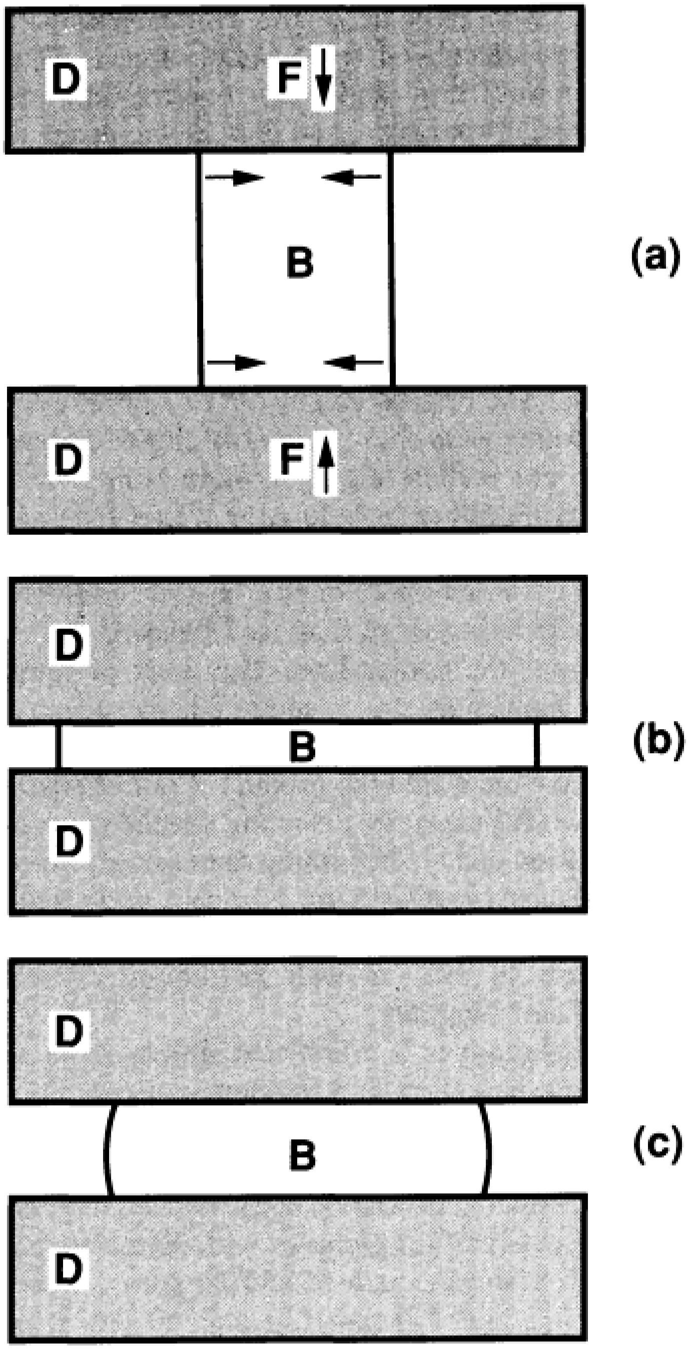


Fig. 8 (a) Billet between two dies. (b) Deformation in the absence of friction. (c) Deformation in the presence of friction. B indicates billet.

Figure 9(a) depicts deformation of strip by a cold rolling process. Metal enters the contact arc at velocity V_1 and exits the contact arc at velocity V_2 and with a reduced thickness. The small arrows along the contact arc denote the direction of frictional force. In rolling, the strip enters the contact arc at a velocity slower than the velocity of the work rolls, and friction is in the direction of metal flow. As strip thickness is reduced in the contact arc, strip velocity increases until it and the work roll velocity are identical. This point is called the neutral point. After the neutral point, further strip reduction occurs, and the velocity of the strip is higher than the velocity of the rolls. Once past the neutral point, frictional force is in the opposite direction of metal flow. Figure 9(b) depicts the pressure distribution within the contact arc in the absence of friction. The area under the curve represents the total pressure required to deform the strip. The pressure generally increases throughout the contact arc, because of strain hardening of the metal, and is therefore higher on the exit side than the entry side. Figure 9(c) illustrates the pressure distribution within the contact arc in the presence of friction.

For the same deformation, the pressure is increased by area A because additional pressure must be exerted to overcome frictional force.

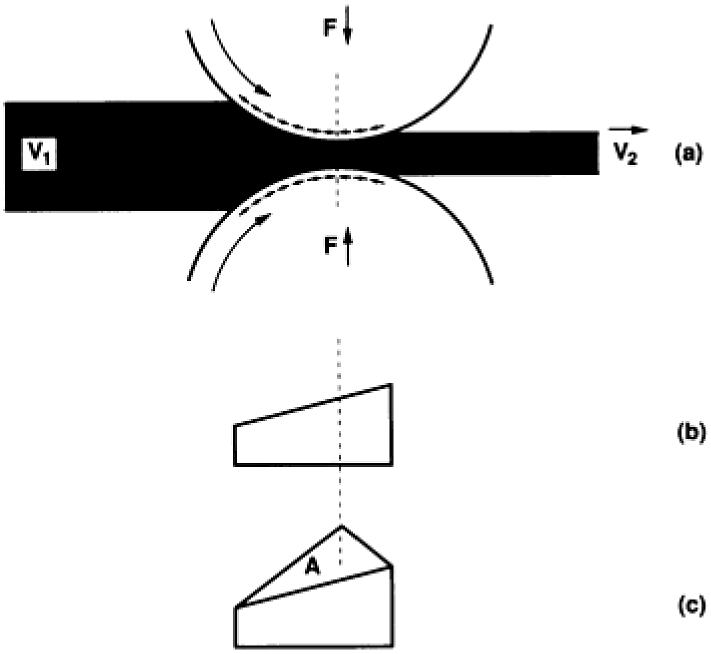


Fig. 9 (a) Deformation of strip by cold rolling. (b) Pressure in contact arc in the absence of friction. (c) Pressure in contact arc in the presence of friction. F is force, A is area.

It should be obvious from these two examples that control of friction in metal forming operations by selection and use of proper lubricants is very important. Friction highly influences power requirements, force requirements, degree of deformation, and operational stability.

Selection of lubricants varies widely depending on the particular metal forming operation and the nature of the workpiece. In the case of ferrous metals, straight oils or emulsions containing fatty materials, as well as extreme pressure additives, are often used. Aluminum metal forming lubricants are commonly straight oils or emulsions compounded with fatty acids, esters, or alcohols. Synergistic blends of these additives are often used. In some cases, such as the body-making operation in the formation of aluminum beverage cans, microemulsions or micellar solutions are successfully used. Lubricants similar to those used for aluminum are successful for copper and copper alloys. Straight oils and emulsions formulated with fatty derived film-strength additives are used. Titanium and its alloys are difficult to form because of the high forces required and a tendency to seize and gall. The surface of titanium is often subjected to surface treatments such as oxidation, nitriding, or cyaniding to lower friction and minimize seizing. Halogenated hydrocarbons, especially fluorocarbons, are effective lubricants for titanium metal forming. Two specific examples of lubricant selection for metal forming are discussed below.

Aluminum Rolling Lubricants. A typical path used to produce commercial aluminum strip begins when a large preheated and scalped ingot enters into a hot reversing mill. The ingot can be several hundred millimeters thick and weigh several thousand kilograms. It is passed back and forth through the reversing mill, often from 15 to 19 times, and its thickness is reduced to about 30 mm (1.2 in.). The sheet is then sent to a hot continuous tandem mill that contains from two to six stands. Here, it is rolled to a gage from 2 to 5 mm (0.08 to 0.2 in.), typically, by employing understand and exit tensions, and then coiled. The coil is typically annealed, allowed to cool, and is then cold rolled on single-stand or multistand mills, under tension, to a gage of about 0.3 to 0.9 mm (12 to 36 mils). Annealing between cold passes to achieve properties is a common practice. The coil is then sent to finishing and prepared for sale.

Because the ingot is not under tension at the reversing mill and large drafts in excess of 30 mm (1.2 in.) per pass are often taken, a major concern is the refusal of the ingot to enter the mill. Reversing mill lubricants are formulated to give high friction to minimize the incidence of refusals, often at the expense of workpiece surface quality. The goal of a reversing mill lubricant formulator is to produce a product that avoids refusals while giving the best possible surface quality.

Typical reversing mill lubricants are very stable emulsions that contain petroleum sulfonate and, often, nonionic emulsifiers with low levels of film-strength additives. The concentrate typically contains oxidation inhibitors, corrosion inhibitors, and antifoaming agents, as well as coupling agents to aid emulsification. Typical concentrate viscosity is about 100 mm^2/s (100 cSt) at 40 °C (104 °F).

At the hot continuous mill, refusals are much less a problem, and surface quality of the workpiece is much more important; therefore, relatively unstable emulsions with more available oil are used. These provide lower friction, but much better surface quality than reversing mill emulsions. Typical lubricants contain petroleum oils and either anionic or non-ionic emulsifiers, and are highly fortified with fatty acid and triglyceride film-strength additives. The concentrate also typically contains the other additives mentioned for reversing mill emulsions and would have a viscosity up to 100 mm^2/s (100 cSt) at 40 °C (104 °F). Because hot continuous mill emulsions are relatively unstable, they have shorter batch lives and are more prone to stability excursions than reversing mill emulsions. Therefore, they must be carefully monitored. The typical percent oil in both hot reversing and continuous mill emulsions is 5%.

Surface quality is of paramount importance in a cold rolling operation. Typical lubricants are straight petroleum oils compounded with small amounts of fatty acids, esters, or alcohols. Typical viscosities range from 3 to 7 mm^2/s (3 to 7 cSt) at 40 °C (104 °F) and are much lower than hot rolling viscosities, because of the lower temperatures.

Rolling oils are maintained through filtration, monitoring of various parameters, and adding components, as required. In the case of the hot mill emulsions, a standard operating procedure for monitoring microbial growth and making biocide additions is also followed.

Steel Warm Forging Lubricants. It is common practice to forge steels over a wide range of temperatures. Cold forging is carried out at ambient temperature, warm forging from about 200 to 700 °C (390 to 1290 °F) and hot forging from about 900 to 1250 °C (1650 to 2280 °F). The choice of temperature employed depends on a balance between sufficient ductility for required formability and the dimensional tolerance required in the forged workpiece. Ductility increases with increasing temperature, whereas dimensional tolerance decreases with increasing temperature. Warm forging often gives an acceptable compromise between ductility and dimensional tolerance.

When forging is performed at room temperature, the billet is commonly subjected to phosphating, in which a zinc phosphate film that aids retention of a soap lubricant is produced. Stainless steels cannot be phosphated, and oxalate films are often used.

At forging temperatures between 400 and 850 °C (750 and 1560 °F), phosphate coatings are ineffective because of oxidation, and are not used. Because molybdenum disulfide begins to oxidize at these temperatures, graphite is the lubricant of choice. Graphite is commonly dispersed in either a water or oil carrier, and is held in suspension by agitation, as well as by either emulsifiers or polymers. Other materials, such as finely divided oxides of tin or lead, can also be present. The lubricant is normally applied to the dies and billet by spraying as a fine mist to ensure complete coating.

Several factors are important for consistent lubrication. The structure, purity, and particle size of the graphite affects results. Large particles have poor film-forming properties, whereas small particles reduce the threshold temperature of graphite oxidation. Particles below about 0.1 μm (4 μin.) become ineffective, because of loss of graphitic structure. Other important factors that require control are the consistency of suspension and the total percent solids.

Handling and Control. It should be evident from the two examples given that lubricant practices are different for dissimilar processes. In the rolling example, the lubricant is reused many times, and issues of handling and control are similar to those already described for metal removal fluids. In the forging example, the lubricant is consumed and destroyed by the process, and the initial composition parameters are controlling factors. The next section in this article briefly addresses modern lubricant practices in terms of quality measures.

Quality

The manufacture of a consistent product requires a consistent process. It should be evident that lubrication is a major factor in metalworking processes, and often plays a dominant role in process consistency. Variability in metalworking lubricants arises from many sources. Incoming lubricant shipments often show batch-to-batch variability in viscosity, additive levels, and interfacial tension, which affects performance. Lubricants in use undergo changes that are due to oxidation, thermal degradation, microbial attack, contamination, and additive depletion.

There are two sources of variation in a process or product. Common sources normally consist of the combined effect of many minor causes and affect the process or product uniformly. Products with variations that are due only to common sources have predictable attributes. Processes with variation that are due to common causes alone produce predictable outputs and are said to be stable. All processes and products have some level of common-source variation.

Special sources of variation affect products and processes in nonuniform ways. They occur in an erratic and unpredictable fashion and are often due to a limited number of significant causes. Products with variations that are due to special causes have unpredictable attributes, whereas variations in processes that are due to special causes yield unpredictable outputs and are therefore unstable, or "out of control."

It has been typical practice to attempt to control metalworking lubricant attributes by setting specifications on incoming supplies and on the lubricants in use and by testing lubricants against these specifications on a regular basis. However, metalworking lubricants can meet all specifications and still be beset by variations that are due to special sources, consequently leading to inconsistent and unstable operations.

Statistical Methods. Statistical process control and statistical quality control use basic statistical tools to detect and reduce special variations in processes and products. Commonly used statistical methods include control charts, histograms, scatter diagrams, and Pareto diagrams. Each is briefly discussed below.

Control charts have many forms, depending on the nature of available data and the parameter to be monitored. A typical control chart, known as an *x*-bar chart (Fig. 10), shows the variation in viscosity of incoming shipments of a rolling oil additive. The chart is based on large samples of measured viscosities for this product. The center line, $\overline{\overline{x}}$, is an estimate of the average viscosity over time calculated from sample means. The upper control limit and lower control limit are calculated statistically such that the incidence of random chance causing the viscosity of a shipment to be outside either limit would occur only 0.3% of the time. When viscosities of shipments lie outside the control limits, the viscosity for individual batches is not predictable, and, by inference, the process used to produce the oil is unstable. The process can be out of control even when points lie within the control limits. Examples would include eight or more consecutive points lying above or below the center line, six or more consecutive points increasing or decreasing in magnitude, or two or three consecutive points lying more than two standard deviations from the center line. The odds of these events happening by random chance are extremely small; therefore, special sources of variation have crept into the product. Control charts are powerful tools for detecting special sources of variation and are becoming increasingly common. Many metalworking lubricant users keep control charts on incoming shipments and on the lubricants in use, as well as request control charts from suppliers.

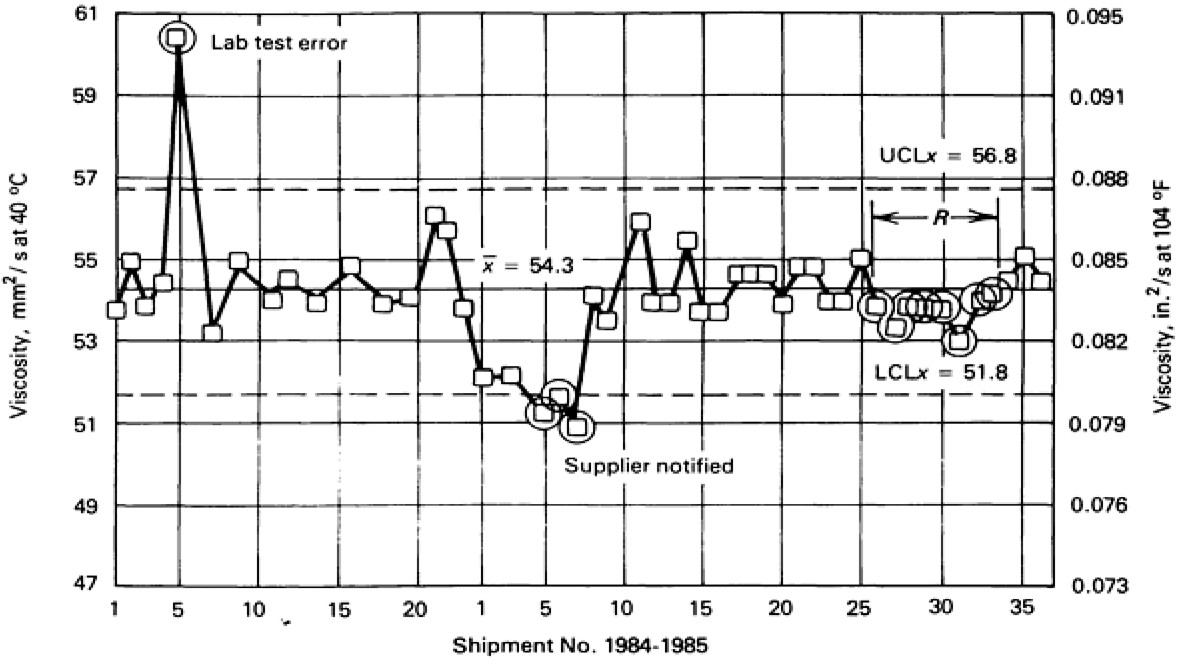


Fig. 10 *x*-bar control chart for viscosity of incoming lubricant shipments

A histogram is a graph that utilizes bars to represent the values in a frequency distribution. In an emulsion, for example, oil globule size can vary between about 0.2 and 10 μm (8 and 400 μin.). A histogram can be constructed that gives the frequency of oil globules with diameters between 0.2 and 0.5 μm (8 and 20 μin.), or between 0.5 and 1.0 μm (20 and 40 μin.), or between 1.0 and 1.5 μm (40 and 60 μin.), for example. With this visual depiction of the frequency of oil globule distribution, a lot of information can be gleaned. It can be determined whether or not the distribution is normal. The average globule size can be determined, along with the spread of the distribution. The difference between the same emulsion under different chemical conditions, pH for example, or at different times also can be determined. Additionally, the variation in oil globule size between emulsions supplied by different sources can be pictured.

A scatter diagram is used to detect the correlation between two variables. If a cause and effect relationship is suspected, then the independent variable is plotted along the abscissa, and the dependent variable is plotted along the ordinate axis. If, for example, observed variation in tool wear in a turning operation was suspected to be due to variation in the concentration of an EP additive, then tool wear would be plotted on the ordinate and EP additive concentration on the abscissa. Often, it is possible to determine if a correlation exists by observing the graph. Statistical methods are available for determining the degree of correlation. Just as control charts are useful in determining whether or not special sources of variation are present, scatter diagrams are useful in determining what those sources are.

Pareto Diagram. In operations where rejections occur, there are generally multiple causes. A Pareto diagram indicates which source of rejections should be solved first. It is really a frequency diagram in the form of a bar chart with the largest source of rejections plotted first, the second largest source plotted second, and so on. Pareto diagrams are made such that the largest source of rejections is plotted on the left, and the smallest, on the right. A lubricant supplier might, for example, want to know the reason for customer complaints and loss of business in a soluble oil line over the past year. Records may show that instability, low viscosity, unacceptable foam, and rust accounted for 100% of the complaints. Further investigation could determine the frequency for each complaint. A Pareto diagram that illustrates that the major cause of complaints was low viscosity, followed by instability, unacceptable foam, and rust in decreasing frequency could be constructed. The diagram visually exposes the relative magnitude of complaints, providing a base of uniform knowledge from which to solve the problem. It also can be used to validate and assess the impact of corrective action.

Total Quality Management. Statistical quality control is one set of tools useful in implementing the larger concept of total quality management or total quality commitment. Other tools include quality function deployment, continuous improvement, statistical experimental design, and problem-solving quality teams. In the future, metalworking lubricant suppliers will have commitment to a continuous and measurable reduction of variation in their processes and products. They also will have a strong focus on all aspects of customer satisfaction. Metalworking lubricant users are beginning to demand these as their customers press them for the same improvements in their products and processes.

Selected References

- D.C. Boley, Application of Basic Statistical Process Control Techniques to Rolling Mill Lubricants, *Lubr. Eng.,* Vol 42 (No. 12), 1986, p 740-750
- A.D. Cron and J. Fatkin, Graphite in Lubrication: Fundamental Parameters and Selection Guide, *NLGI Spokesman,* Vol 53 (No. 4), 1989, p 137-147
- M. Fukuda, T. Nishimura, and Y. Moriguchi, Friction and Lubrication in the Fabrication of Titanium and Its Alloys, *Metalwork. Interfaces,* Vol 5 (No. 3), 1980, p 14-21
- K. Glossop, Copper Wire Drawing Lubricants, *Wire Ind.,* Vol 56 (No. 661), 1989, p 45-49
- "The HLB System," ICI United States, Inc., 1976
- K. Ishikawa, *Guide to Quality Control,* 2nd ed. (revised), Asian Productivity Center, Tokyo, 1986
- C. Kajdas, Additives for Metalworking Lubricants--A Review, *Lubr. Sci.,* Vol 1 (No. 4), 1989, p 385-409
- S. Kalpakjian, Recent Progress in Metal Forming Tribology, *J. Appl. Metalwork.,* Vol 4 (No. 3), 1986, p 270-280
- R. Klann, "Harmless" Lubricants and Coolants Can Poison Your Carbide Tooling, *Met. Stamp.,* 1979, p 3-5
- I. Mizutani and I. Iwasaki, Recent Trends in Precision Forging, *JSAE Rev.,* Vol 9 (No. 3), 1988, p 70-75

- M. Mrozek, Recent Developments in Solid Lubricants for Metalworking: Part 1--Graphite, *Metalwork. Interfaces,* Vol 4 (No. 1), 1979, p 13-22
- E.S. Nachtman and S. Kalpakjian, *Lubricants and Lubrication in Metalworking Operations,* Marcel Dekker, 1985
- L.M. Prince, Ed., *Microemulsions Theory and Practice,* Academic Press, Inc., 1977
- J. Saga, Lubrication in Cold and Warm Forging of Steels, *Metalwork. Interfaces,* Vol 5 (No. 1), 1980, p 9-23
- J.A. Schey, *Tribology in Metalworking: Friction, Lubrication and Wear,* American Society for Metals, 1983
- H.E. Sliney, Solid Lubricant Materials for High Temperatures--A Review, *Tribology International,* 1982, p 303-315
- M. Suzuki, Lubrication in the Rolling of Aluminum, *Metalwork. Interfaces,* Vol 5 (No. 2), 1980, p 10-22
- S.K. Tinker, The Use of Quality Methods in the Production of Linear Paraffins, *Lubr. Eng.,* Vol 46 (No. 11), 1990, p 697-703
- D. Zintak, Ed., *Improving Production with Coolants and Lubricants,* 1st ed., Society of Manufacturing Engineers, 1982

Lubricants for High-Vacuum Applications

Michael R. Hilton and Paul D. Fleischauer, Mechanics and Materials Technology Center, The Aerospace Corporation

Introduction

THE SELECTION CRITERIA and performance of lubricants in the vacuum environments that exist in terrestrial equipment and both in and around space vehicles are described in this article, as are the vacuum, mechanical, and thermal conditions of such lubricant use. The tribology issues that should be considered in the initial design of systems used in terrestrial equipment or space vehicles are also defined. The types of vacuum lubricants that are available (dry or solid, liquid, or grease) are reviewed, their properties and applicability to general situations are assessed, and the option of not lubricating is discussed. Examples of specific applications and performance data for lubricants in space and on earth are presented, as well.

Scope of the Problem

The pressures (vacuum) to be considered for lubricant performance range from 1.3×10^{-2} Pa to 1.3×10^{-10} Pa (1×10^{-4} to 1×10^{-12} torr). This range encompasses vacuum systems used for thin-film deposition or materials processing (base pressure of 1.3×10^{-4} Pa, or 1×10^{-6} torr, or higher) and for surface science experiments (1.3×10^{-9} Pa, or 1×10^{-11} torr, or higher). Although deep-space vacuum can reach the 10^{-12} Pa (10^{-14} torr) range, near-earth orbits have higher pressures, perhaps 1.3×10^{-6} Pa (1×10^{-8} torr). Continual outgassing within a space vehicle is estimated to expose any internal mechanism to a pressure of 1.3×10^{-5} Pa (1×10^{-7} torr) or lower (Ref 1). Externally exposed mechanisms in low-earth orbit (less than 483 km, or 300 miles) may experience atomic oxygen bombardment with an apparent flux of 10^{7} to 10^{16} cm^2/s (Ref 2).

A potential problem created by these vacuum levels is that conventional liquid lubricants, which usually have relatively high vapor pressures ($\geq 1.3 \times 10^{-4}$ Pa, or 1×10^{-6} torr, at room temperature) and surface diffusion coefficients ($\geq 1 \times 10^{-8}$ cm^2/s, or 1.5×10^{-9} $in.^2/s$) with low surface tensions ($\sim$0.018 to 0.030 N/m, or 18 to 30 dyne/cm), can either volatilize or creep away from the area of mechanical contact. The result is high friction, wear, or mechanical seizure.

The volatility problem can cause lubricant vapor pressure to limit the achievable vacuum baseline of the system. Further, this problem can cause the lubricant to migrate to and condense on (that is, contaminate) sensitive surfaces, such as solar cells, optics, or the material to be probed in a surface science experiment. However, the volatility problems of liquid lubricants can be circumvented by design features that include proper confinement geometries, or by using new synthetic oils that have extremely low vapor pressures, or both. These two solutions are reviewed in the section "Types of Vacuum Lubricants" in this article.

Another potential tribological problem created by vacuum stems from the removal of reactive gases, particularly water vapor, oxygen, and some carbonaceous species, that are present in the atmospheric environment. Normally, these reactive gases chemically passivate the near-surface region (1 to 5 nm, or 0.04 to 0.2 μin.) of most materials, especially metals, significantly inhibiting the welding (adhesion) of surfaces upon contact. These passive layers are often brittle and are worn away during mechanical contact.

The presence of reactive gases in the atmosphere continually repassivates the material, but, in vacuum, such repassivation can be inhibited or eliminated. Therefore, the force of adhesion between freshly exposed metal surfaces will be quite strong upon contact. The joined areas will only separate by fracture, generally accompanied by material transfer from one surface to another. This process of adhesive wear results in the consumption of extra power to drive the mechanism (with possible total prevention of any motion) and degrades the dimensional tolerance of the mechanism components, causing mechanical noise (for example, torque hash, vibration) in precision systems. Thus, contacting metal surfaces that might not require lubrication in atmospheric conditions may require antiseize lubricants to prevent cold-welding in vacuum.

Ceramics also can be lubricated by carbon-containing reactive gases. However, along with semiconductors and polymers, ceramics are not as susceptible to the "cold-welding" phenomenon as are metals.

There are two categories of systems in which ultrahigh-vacuum lubricants are needed: systems in terrestrial vacuum chambers and systems in space vehicles. The first category includes chambers used for surface science experiments; analytical instruments, such as electron microscopes; thermal vacuum testing chambers; thin-film deposition chambers; and other materials processing equipment primarily used in the semiconductor industry. Such systems often have manipulators to move objects within the most critical vacuum region. Contamination by the lubricant is a prime issue for terrestrial vacuum systems, whereas lifetime is a lesser concern, because the lubricant can be periodically replaced by breaking vacuum.

In space vehicle systems, human intervention is essentially impossible (except infrequently, when a satellite in low-earth orbit can be retrieved by a manned shuttle). Therefore, both lifetime and contamination are of great concern.

Although the conditions of mechanical contact in vacuum systems are varied, some generalizations can be made. Because most mechanisms do not operate continuously, there are periods of boundary contact (when lubricant film thickness is less than the contacting surface roughness) between component surfaces. Similarly, boundary contact also develops when oscillating mechanisms change direction. In such situations, the chemical interaction of contacting surfaces, often modified by the presence of lubricants, is critically important.

Both sliding and rolling elements are designed into vacuum systems. Each element requires proper consideration of lubricant specification to prevent cold-welding, to promote low-torque noise performance, and to ensure adequate service life. Other mechanical parameters that are important in the design of moving assemblies and in the selection of the proper lubricant include the expected loads and contact stresses, the geometry of the contact (the conformity of the surfaces and the possibility of lubricant confinement), and the relative velocities (rotational speed) of the contacting surfaces.

Temperatures encountered by vacuum components usually range from 0 to 75 °C (32 to 165 °F), although mechanisms within infrared sensors are cryogenically cooled. Lubricants that function above 200 °C (390 °F) are beyond the scope of this article. Interested readers are referred to other sources (Ref 3, 4, 5).

Another possible condition for lubricants in vacuum is the requirement for electrical conductivity. Some satellites maintain attitude stability by spinning the entire vehicle while the antennas and sensors are continually despun to allow them to remain pointed at fixed objects. Other satellites have rotating panels of solar cells that track the sun to maintain constant power and to keep batteries charged. Electrical signals and power must be transmitted across a rotating interface, which consists of sliding wipers that have a conductive lubricant. This issue will be explored later in the text.

Yet another condition is the possible use of high-speed bearings in satellite systems, such as those in fly wheels used for momentum stabilization and in turbomachinery (rocket motors). Because of frictional heating in such bearings, lubricants (usually oils or greases) are required that can conduct heat efficiently among rolling elements and maintain low overall operating temperatures. (The unusual case of liquid hydrogen and liquid oxygen pumps, as used in the space shuttle main engines, involves solid lubrication of the bearings with fluid, or fuel, coolants. The involvement of the fluids in lubrication is probably minimal).

Ideal Tribological Situations and Considerations

Tribology often does not receive sufficient and timely consideration during the design phase of a mechanical system that is to function in vacuum. Such inattention can convert an expensively fabricated mechanism into scrap material, increasing system cost and causing schedule delays. This section of this article reviews the ideal or desirable characteristics of tribological contact in vacuum, as well as the general principles involved in achieving these characteristics.

The ideal characteristics for lubricants in vacuum are:

- Low vapor pressures
- Acceptable lubricant creep or migration (including little or no lubricant debris formation)
- Long life (meeting system service life with margin)
- Low friction (including low power consumption, low heat generation, low disturbances, and low torque noise)
- No wear/no significant deformation
- Temperature insensitivity
- Suitable electrical conductivity

Low Vapor Pressure. As stated previously, low vapor pressures are desirable to prevent lubricant loss away from the region of mechanical contact, which can lead to mechanism seizure, and to prevent contamination of critical surfaces. Conventional mineral oils, even those that are super refined, or molecularly distilled, consist of a broad distribution of molecular species. The lighter-weight members of the distribution, which can be significant in number, are volatile and result in a moderately high vapor pressure ($\sim$1.3 × 10^{-4} Pa, or 1 × 10^{-6} torr) for the oil. Generally, 20 to 30% of such an oil will evaporate before the vapor pressure drops significantly (Ref 6).

Conversely, synthetic hydrocarbon oils, such as poly-α-olefins, neopentyl polyolesters, and other tailored polymers, are made with very narrow distributions of molecular weights. Therefore, both vapor pressure and viscosity are controlled to give optimum values. Less than 3% of these oils evaporates in a high-vacuum chamber at elevated temperatures (pressure $\geq$1.3 × 10^{-4} Pa, or 1 × 10^{-6} torr, at 100 °C, or 212 °F).

Additives included in oils to provide antiwear or antioxidation protection usually have higher evaporation rates and vapor pressures than the base oils themselves. Such differences can cause problems, because if the additives evaporate, they can become sources of contamination and deplete the base oil of protection. Solid lubricants generally have negligible vapor pressures relative to liquid lubricants.

Acceptable Lubricant Creep or Migration. Another phenomenon that can cause problems similar to those of high vapor pressures is surface diffusion, or creep, of the liquid lubricant away from the contact region. Creep is associated with characteristically low (0.018 to 0.030 N/m, or 18 to 30 dyne/cm) surface tension of the lubricant on the component surface. Such low surface tension is desirable, because it promotes wetting. Therefore, creep is generally countered at the system design level by including antimigration barriers, as discussed in the section "Types of Vacuum Lubricants."

Extremely low surface tension ($\sim$0.018 N/m, or 18 dyne/cm), and, therefore, significant creep problems are encountered with synthetic perfluorinated polyalkylether oils. These oils have been used extensively in spacecraft because they can have even lower vapor pressures than the above-mentioned synthetic hydrocarbons. However, the hydrocarbons are preferred in many applications, because problems with creep are considerably reduced and they can be formulated with additives to produce far superior lubricants. Although not in common use at present, another class of oils known as silahydrocarbons (Ref 7, 8) offers potential advances for vacuum uses.

For solid lubricants, there is a similar concern about the presence and migration of particles generated from detached film debris. This possible problem is not well documented. Therefore, line-of-sight barriers and the effects of gravity should be considered to contain particle migration. In space, particulate motion may be exacerbated by the lack of gravity. On earth, prudent designers should locate mechanisms below critical surfaces or operations, allowing gravity to pull particles away harmlessly.

Long Life. A lubricant must have a long operational life to be considered successful. However, long life in this sense is relative to the anticipated service life of the mechanism. The point is that a lubricant should be chosen that has an adequate endurance life (including a reasonable safety margin) and not necessarily the best endurance life, because other performance properties and design issues also have to be considered.

Low Friction and Wear. Low friction is important to vacuum mechanisms, in order to reduce the consumption of power, which is supplied to spacecraft by batteries and solar cells, and is therefore finite. Low friction also reduces heat generation. For mechanisms that are controlled either electronically or in a feedback loop, a stable ("hash-free") friction coefficient (low noise) is particularly important to maintain proper control. In spacecraft, attitude control is maintained by momentum transfer mechanisms (control moment gyroscopes, momentum wheels, or reaction wheels) or by spinning most of the vehicle. Variable friction of the bearings at the despin interface can cause the vehicle to wobble or tumble.

Lubricants that promote low friction also retard wear and the plastic deformation of contacting surfaces. Such wear and deformation can lead to loss of component tolerance, which, in turn, can cause increased torque noise and/or variable torque levels, or outright mechanism failure. At the atomic level, reducing friction is synonymous with minimizing chemical bonding between contacting surfaces.

There is a positive relationship between friction and chemical reactivity; the friction coefficient of metals against themselves in vacuum correlates with the position of the metal on the periodic table (Ref 9). For example, the d-shell metals on the left side of the table, which are more reactive, have been found to have higher friction coefficients. Thus, it is desirable for the surfaces of contacting components to have low chemical reactivity. The lubricant modifiers surface composition to achieve this reduced reactivity. In unlubricated situations, low surface reactivity favors the selection of either polymers or ceramics.

Ceramics are particularly interesting because their bonding properties not only yield lower surface reactivity, but higher elastic modulus and strength characteristics. The latter two properties help resist deformation and irreversible loss of tolerance. The designer should only select component materials that have elastic limits greater than the operational stresses of the mechanism to avoid plastic deformation or fracture and subsequent loss of component tolerance.

The ideal tribological contact can also be viewed from the macroscopic continuum perspective. For example, the friction coefficient between unlubricated or dry lubricated surfaces is (Ref 10):

$$\mu = (\tau_s)\,(A)/W \qquad \textbf{(Eq 1)}$$

where μ is the friction coefficient, τ_s is shear strength of the weakest material or interface, A is the true contact area, and W is the normal contact load. The τ_s is related to kinetic friction force, F_k, by:

$$F_k = (\tau_s)\,(A) \qquad \textbf{(Eq 2)}$$

For a Hertzian contact in which a smooth sphere is sliding against a flat surface, Eq 1 can be rewritten (Ref 11):

$$\mu = (s/W)\,(3R/4E)^{2/3} \qquad \textbf{(Eq 3)}$$

where R is the radius of the contacting sphere and E is the effective modulus of the contacting materials. For a given load, the use of a lubricant or surface modification process that reduces the shear strength of the interface will reduce friction. If the lubricant film or surface-modified region is thin, the load is supported primarily by the substrate. Increasing the substrate modulus decreases the contact area for a given load, which also reduces friction.

The designer can vary the surface composition of components either directly or by placing lubricants on the contacting surfaces, as will be discussed in the section "Types of Vacuum Lubricants" in this article. Component sliding can, in

principle, occur by the shear or relative motion of atoms along only two adjacent contacting lattice planes. Thus, the above atomistic and macroscopic arguments mean that the ideal contact consists of a thin surface zone (including the lubricant, if present, and component surfaces) having low shear strength (or low reactivity), with the bulk material(s) of the components providing a high-modulus, underlying support. The ideal surface zone or lubricant film thickness has been found to be between 0.5 and 1.0 μm (20 and 40 μin.) (Ref 11).

Temperature Insensitivity. Lubricants must provide acceptable friction performance within the operational temperature limits of the system. Solid lubricants generally have less temperature sensitivity than liquids and greases. Again, the designer must remember to consider all lubricants that have adequate temperature characteristics for the given application, and not necessarily to choose the lubricant with the best temperature characteristics. The thermal conductivity characteristics must also be viewed with similar criteria.

Suitable Electrical Conductivity. For most mechanical contacts in vacuum, electrical conductivity across the contacting surfaces and through the lubricant is not an issue. However, conductive lubricants are required in sliding electrical contacts found in many space systems. These contacts transmit power and signal between different spacecraft sections that move relative to each other, that is, motion often occurs between the main body of the spacecraft and other subsystems, such as solar cell arrays, antennas, and sensors. As will be discussed below, solid-lubricant composites are frequently used.

Types of Vacuum Lubricants

Three types of lubricants are used in vacuum environments: solid (dry), liquid, and grease. There is also a fourth approach, which is to use no lubricant at all, but to rely instead on the low reactivity of the contacting surfaces. The component materials have to be carefully chosen and, perhaps, have their surface compositions modified to lower reactivity. Descriptions of most solid and liquid space lubricants, as well as conditions for use, are given in Ref 12, which is a comprehensive treatise that should be consulted before choosing a vacuum lubricants. This section defines and reviews the available lubricants, and assesses their favorable and unfavorable properties. Whenever appropriate, methods of application or processing are included.

Solid (Dry) Lubricants. There are four types of solid, or dry, lubricants available for vacuum applications: soft metals, lamellar solids, polymers, or other soft solids (Table 1). Composites of these four lubricant types or combinations of one or more of them with matrix or support materials are also available.

Table 1 Potential solid lubricants for use in vacuum

Soft metal films
Au, Ag, Pb, In, Ba
Lamellar solids
Dichalcogenides (MoS_2, WS_2, $MoSe_2$)
Pthalocyanines
$CdCl_2$, $PbCl_2$
Intercalated graphite
Boron nitride
Polymers
PTFE, FEP, polyacetal, polyimide
PEEK, UHWPE
Phenolic and epoxy resins
Other low shear strength solids
Oxides: Cd, Co, Zn
Sulfides: Bi, Cd
Fluorides: Ca, Li, Ba, rare earths

Source: Ref 13

The major advantages of solid lubricants, as indicated in Table 2, are that they have negligible vapor pressures and some are quite insensitive to temperature. Therefore, they are the only lubricants suitable for cryogenic and high-temperature applications. Accelerated testing is valid for solid lubricants if the same wear mechanism is operative at different test

speeds (which can be verified by microscopy). Because space missions often exceed 5 years, accelerated ground testing becomes essential for the qualification of new lubricants. Generally, solid lubricants have lower friction coefficients in vacuum than greases, and can have lower friction coefficients than liquids. For example, MoS_2 films have been developed with friction coefficients lower than 0.1. Although Teflon has been listed as the material with the lowest friction coefficient known, at 0.02 (Ref 14), MoS_2 films can have considerably lower values, as low as 0.007 under the right conditions (Ref 11).

Table 2 Relative merits of solid and liquid lubricants for use in vacuum

Dry lubricants
Negligible vapor pressure
Wide operating temperature
Negligible surface migration (debris can float free)
Valid accelerated testing
Short life in laboratory air[(a)]
Debris causes frictional noise
Friction speed independent
Life determined by lubricant wear
Poor thermal characteristics
Electrically conductive
Wet lubricants
Finite vapor pressure
Viscosity, creep, and vapor pressure all temperature dependent
Seals required
Invalid accelerated testing
Insensitive to air or vacuum
Low frictional noise
Friction speed dependent
Life determined by lubricant degradation
"High" thermal conductance
Electrically insulating

Source: Ref 13

(a) Depends on type and matrix; for example, some polymers and bonded solids, especially graphite, behave well in air.

The major disadvantage of solid lubricants is their shorter lifetime relative to liquids or greases. Once a solid lubricant is pushed out of the contact zone, lubricant resupply generally does not occur (except for transfer film schemes, as described shortly), as it does for liquids or greases. Also, failure is quite abrupt (catastrophic, compared to graceful failure observed for most fluid lubricants), often with no prior performance degradation visible. The solid lubricant that is pushed out of the contact zone forms debris, which can, in turn, form bumps and lead to torque disturbances in precision bearings or become unwanted particulate material.

Solid lubricants can be applied by rubbing (burnishing) a power or a solid block of lubricant against a component surface, resulting in transfer of the lubricant to the critical surface, or by applying the lubricant as a thin film to the component prior to mechanism use. The rubbing approach can be used to develop a source of lubricant resupply if some portion of the mechanism is fabricated from the lubricant. Ball-bearing cages (retainers) made of polymer-based composites or of leaded bronze have been used in this way. The disadvantage of the rubbing approach is that lubricant transfer can be sporadic or nonuniform, yielding lubricant bumps or bare regions on the contacting surfaces. For precision mechanisms, unacceptable torque noise can result.

Solid-lubricant films can be applied by either rubbing or burnishing, although careful procedures have to be followed to maximize even coverage. Another alternative that is used for lamellar compounds is to mix the lubricant with a binder and a solvent and apply the mixture by dipping, painting, or spraying. The resulting bonded films often require air or heat curing after application. Bonded films are generally several micrometers thick, which often does not allow for the lowest possible friction of low shear strength materials, and which is dimensionally unsuitable (too thick) for many precision components. However, the bonded-film technology is well established and is quiteeffective and appropriate for many low-cycle applications that cannot tolerate seizure (Ref 12, 15).

Another approach to applying thin films of dry lubricants is to use vacuum deposition techniques for uniform coverage of components in precision systems. When film thickness is less than 1 μm (40 μin.), low and steady friction is obtained, as shown in Fig. 1 (Ref 16). For precision mechanisms, the films can be applied by sputtering, ion plating, or other ion beam assisted techniques to obtain even, controlled lubricant coverage. Ion cleaning of the substrates prior to deposition can be used to improve film adhesion to the components. A major disadvantage of solid lubricant thin films is that there is no means of lubricant resupply. Therefore, lubricant endurance life relative to system service life is of prime importance when considering solid lubricants.

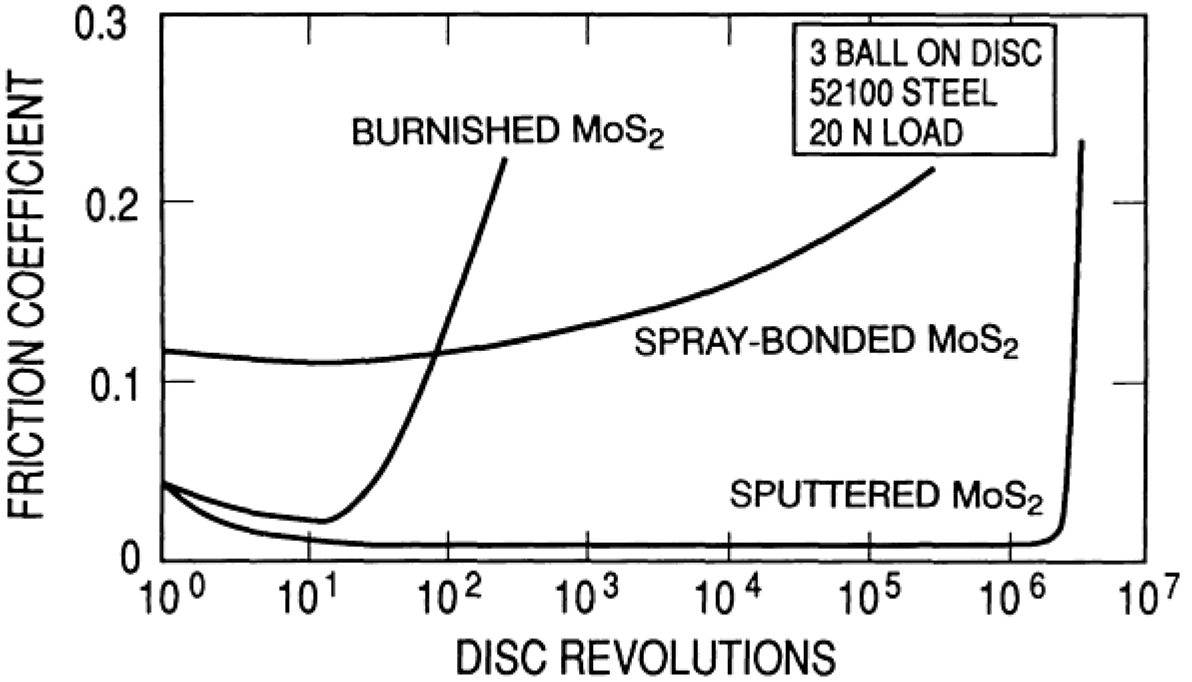


Fig. 1 Friction coefficient of various types of MoS_2 as a function of cycles in pin-on-disk tests. Source: Ref 16

Soft metals, including lead, gold, silver, and indium, have all been used as lubricants in vacuum applications (Ref 17). Of these metals, lead has had the most success and use. Burnishing and electroplating have been used to apply lead. However, deposition by ion plating provides the best adhesion and is preferred for uniform coverage. Optimum performance of lead and other metals is achieved at approximately 1 μm (40 μin.) thickness. Ion-plated lead films have been particularly effective in spacecraft bearings found is solar array drive mechanisms, especially in European satellites. Silver and gold are useful in situations requiring electrical conductivity. However, silver is generally too hard for most applications, and gold work-hardens quite easily. Lead remains soft at room temperatures, and evidence indicates that it can lubricate at 20 K (Ref 17).

Lamellar solids that are in relatively wide use as lubricants include the disulfides and diselenides of Mo, W, Nb, and Ta. Graphite is also a lamellar solid lubricant, but the pure material is not suitable for vacuum applications, as will be discussed below (Ref 5, 15). Some other doped or intercalated layered solids have been investigated for their lubrication properties, but they are not in wide use (Ref 18). The anisotropic, planar crystal structures of lamellar solids provide low-shear planes for lubrication. These solids also have high load-bearing capacity when compressed in a direction perpendicular to their low-shear planes. This load-bearing capability is an advantage of lamellar solids, compared to solid polymer lubricants.

Of the lamellar solids, MoS_2 films deposited by sputtering have been the most widely investigated and developed, since early in the space program (Ref 19, 20) and especially since the late 1980s (Ref 11, 21, 22, 23, 24, 25). MoS_2 films have a lower friction coefficient than lead films ($\leq$0.01 versus 0.1 in vacuum, respectively), which lowers mechanism torque and power consumption (always a concern on spacecraft). MoS_2 films are also superior to lead films in pure sliding applications. Sputter-deposited MoS_2 has superior endurance and a lower running friction coefficient than either burnished or bonded MoS_2, as shown in Fig. 1.

The performance of sputter-deposited MoS_2 is critically dependent on film microstructure, which includes composition, morphology, crystallinity, and preferred orientation (Ref 19, 26). These properties, in turn, are very dependent on deposition conditions; the presence of water vapor during deposition is a particularly insidious variable (Ref 27). The general trend in film development in recent years has been the production of dense films with low porosity, because porosity leads to large-scale film debris generation early in wear (Ref 26). Most films grow with their low-shear basal planes perpendicular to the substrate. Reorientation of the basal planes to a parallel alignment with the substrate occurs during wear.

Stress-induced crystallization has also been observed after sliding wear in some dense films that were disordered as-deposited (Ref 28). There are several deposition practices that can yield these dense films, including high growth rates (Ref 23), low deposition pressures (Ref 24), ion bombardment during film growth (Ref 29, 30), and the incorporation of dopants (Au, Ni, water vapor) that are either co-sputtered continuously (Ref 26, 27, 31, 32) or deposited as multilayers (Ref 33). Some of these films have an initial preferred orientation of low-shear basal planes parallel to the substrate.

MoS_2 is very sensitive to water vapor, although not as sensitive as polyimides, which are discussed in the next section. If MoS_2-lubricated components are stored in a humid environment, significant oxidation will occur over months, forming MoO_3, which is an inferior lubricant (Ref 34). This storage problem is especially relevant for satellites (and vacuum mechanisms) that are assembled at least a year before launch (use). Satellites containing MoS_2 have to be stored in dry, inert-gas environments until shortly before launch.

In fact, there are often several environmentally sensitive materials on satellites that mandate controlled storage. However, recently developed MoS_2 films that have dense morphologies may have better storage oxidation resistance, although no oxidation data are currently available. Such storage has the added benefit of better maintenance of vehicle cleanliness, although the moisture issue can cause some contention, because electrical systems often prefer a moderate relative humidity to prevent static electrical discharges.

MoS_2 does not lubricate as well in a humid environment as in vacuum, where friction coefficients decrease and endurance increases (Ref 15). Indeed, MoS_2 performs at its best in vacuum. [If a MoS_2-coated components rests in vacuum, then water vapor will deposit and slightly oxidize the top surface over time. The extent of oxidation depends on the vacuum level. An initially higher friction will be observed. However, the oxidized layer is quickly removed by either frictional heating and volatilization or by being pushed aside. The underlying MoS_2 exhibits a lower friction coefficient than the top surfaces (Ref 17, 25)].

With regard to water vapor effects, MoS_2 is a direct complement to graphite, which is an excellent lubricant in atmospheric environments. However, the low shear strength of graphite is critically dependent on the intercalation of absorbed gases, especially water vapor (Ref 35, 36). At pressure below $\sim 10^{-2}$ Pa (10^{-4} torr), such gases desorb from graphite, and its friction coefficient dramatically rises. Intercalated graphite compounds that work well in vacuum have been developed (Ref 5). These compounds are not widely available and so far have only been applied by burnishing or used in bonded films (Ref 37).

Although other disulfides and diselenides have been considered for vacuum applications, none of them have the endurance of MoS_2. Although, MoS_2 is a semiconductor, it is routinely used for sliding electrical contacts. $NbSe_2$, which is a semimetal in its natural state, should, in principle, be better for this application. Unfortunately, $NbSe_2$ is not a good lubricant in semimetallic form. It becomes a good lubricant only when it is intercalated with electron donor atoms (which could be excess Nb atoms), whereupon it also becomes a semiconductor (Ref 22).

A widely used additive for liquid lubricants in automotive applications is WS_2, which also performs well at higher temperatures and is more oxidation resistant than MoS_2. Although WS_2 has been deposited by impact under high air pressure in one commercial process, experience has shown that it should be considered primarily as an antiseize compound for limited-use (low-cycle) vacuum applications. Sputter-deposited films of WS_2 or WSe_2 codeposited with MoS_2 can have properties comparable to, or even better than, sputtered MoS_2 films (Ref 38, 39). However, MoS_2 is much more common and generally has superior endurance in vacuum applications.

Polymers and Polymer Composites. Polymers, consisting of anisotropically bonded (highly linear) molecules, can provide low friction surfaces in vacuum, if the molecule chains align properly at the contacting surface. However, because the load-bearing capability of polymers is generally low, additives are required to strengthen the polymer to avoid ploughing into the bulk.

For vacuum applications, polymer composites, rather than pure polymers, are generally used (Ref 40, 41). Because these composites have structural integrity, self-lubricating composite components can be fabricated that can, in principle, provide a continual source of solid lubricant to critical components.

To date, polytetrafluoroethylene (PTFE) has been the polymer used the most in vacuum. This is because PTFE performs well in vacuum and in the presence of absorbed vapors. However, PTFE has a tendency to cold-flow under load, necessitating a binder to restrain the polymer bulk, that is, to prevent ploughing. Some polyimides appear to be excellent in vacuum because they exhibit low friction coefficients without significant cold-flowing of the bulk (Ref 42). However, polyimides are very sensitive to water vapor absorption. Water molecules appear to hydrogen bond to the polymer molecules and then inhibit molecular shear. Thermal pretreatment of polyimides appears to be essential for good performance in vacuum.

Polymer composites include other materials that are added for several reasons: to increase load-carrying capacity, to lower the friction coefficient and promote a low wear rate, and to increase thermal conductivity. Table 3 lists polymers and additives that can be included in self-lubricating composites. Both fibers and particulate additives can be used, although fibers are more effective for increasing composite load-carrying capacity. Studies indicate that, in some composites. MoS_2 facilities polymer transfer to a critical component; the polymer is the primary lubricant, not the MoS_2 (Ref 22, 43).

Table 3 Plastics and fillers for self-lubricating composites

Material	Maximum useful temperature	
	°C	°F
Thermoplastics		
Polyethylene (high MW and UHMW)	80	175
Polyacetal (homo- and co-polymer)	125	255
Nylons (types 6, 6.6, 11)	130	265
Poly (phenylene sulphide)	~200	390
Poly (tetrafluoroethylene)	275	525
Poly (*p*-oxybenzoate)	300	570
Thermosetting		
Phenolics	~150	300
Cresylics	~150	300
Epoxies	~200	390
Silicones	~250	480
Polyimides	~300	570
Reinforcements		
Glass fibers	. . .	
Asbestos fibers	. . .	
Textiles (polyester, "Nomex," cotton)	. . .	
Mica	. . .	
Friction and wear-reducing additives		
Graphite	. . .	
Molybdenum disulphide	. . .	
Polytetrafluoroethylene (PTFE)	. . .	
Metal oxides	. . .	
Silicon fluids	. . .	
Thermal conductivity adjunctives		
Bronze	. . .	
Graphite	. . .	
Silver	. . .	

Source: Ref 41

Table 4 lists some available self-lubricating composites and their possible uses in space. When a polymer composite is selected for an application, contact stress is probably the most important property to consider. The ideal composite must

support, in the bulk of the material, the dynamic stresses of the application, and must allow for the formation, by local deformation, of a low shear strength layer at the surface.

Table 4 Self-lubricating composites and possible uses in space

Composite type	Use
PTFE/glass fiber	Bearing cages
PTFE/glass fiber/MoS_2	Bearing cages, gears
Ployacetal homopolymer/co-polymer	Bearing cages, gears, bushings, brakes
Reinforced phenolics	Bearing cages, gears
Polyimide/MoS_2	Bearing cages, gears
PTFE/woven glass fiber/resin	Bushings
PTFE/bronze sinter	Bushings, rotating nuts

Source: Ref 41

Nonpolymer-based Composites. Two examples of nonpolymer-based composites are particularly worthy of attention. Leaded bronze composites, as one example, have been fabricated into bearing retainers (Ref 11, 17). When used in conjunction with lead-coated bearings (for example, in solar array drive mechanisms), the lead in the retainers provides an effective supplemental source of lead when the original film is worn. Additionally, composite blocks of silver, MoS_2, and either graphite (Ref 44) or copper (Ref 45, 46) are used as brushes in sliding electrical contacts. The silver provides conductivity and structural integrity, the MoS_2 lubricates in vacuum, and the graphite or copper may lubricate in air.

Liquid Lubricants. Examination of Table 2 would suggest that the merits of solid lubricants frequently exceed those of liquid lubricants in vacuum applications. However, liquid lubricants are often used in space applications, particularly on U.S. systems. In the early years of the space program, liquid lubricants were understood better than solid lubricants. Thus, mechanisms were engineered to make low vapor pressure liquids work in vacuum applications (Ref 36, 47). Early success with liquids slowed the incorporation of solid lubricants in U.S. space systems. In contrast, European-designed space systems have often incorporated solid-lubricant technology as it has evolved (Ref 13).

The primary advantage of liquid lubricants over solid lubricants is their long life in high-cycle applications, such as in gyroscopes. Long life results because liquid lubricants can be resupplied and they have low frictional noise in bearing applications. Another advantage is that liquid lubricants have high thermal conductance, which may assist in managing thermal stability on spacecraft. However, for terrestrial vacuum systems, these advantages often do not outweigh the disadvantage of potential contamination from fluids. Therefore, in terrestrial vacuum systems, the use of modified surfaces and/or lubrication by solids or greases is the preferred approach for manipulators and other mechanisms.

A particular disadvantage of liquid lubricants for space applications is that accelerated testing, while desirable, is difficult because of the synergistic dependence of lubricant properties, such as film thickness, on operational parameters, such as contact speed, load, and temperature. It is reasonable to compare the performances of two or more different fluid lubricants in tests in which operational conditions are intentionally more severe than for the expected application. But extreme care is required in the selection of the parameters to be accelerated, and specific acceleration factors should never be quoted for an application without complete theoretical justification. Such justification would have to involve a rigidly verified, mechano-chemical model for determining operational life of the system (or component) of interest.

For space applications, contamination and lubricant loss are minimized by proper lubricant selection and mechanism design. Low vapor pressure lubricants do exist, as shown in Table 5 and Fig. 2. Their lubrication properties will be discussed subsequently. If a component is not entirely sealed, lubricant loss by vapor transport is generally diminished by incorporating molecular seals, often of a labyrinth geometry (Fig. 3), into the bearing design (Ref 44, 47). Contamination, via vapor transport, of critical surfaces away from the tribocontacts can also be avoided by the use of vents into space, pointed away from critical surfaces and away from the leading (front) edge of the vehicle.

Table 5 Properties of selected fluid lubricants

Property	PFPEs			PAOs		POEs (NPEs)		MAC	SiHCs	
	Demnum S200	**Fomblin Z25**	**Krytox 16256**	**Nye 179**	**Nye 188B**	**Nye UC7**	**Nye UC9**	**Pennzane SHF X2000**	**SiHC1**	**SiHC2**
Average molecular weight	8400	9500	11,000						1480	1704
Kinetic viscosity at 20 °C (70 °F) 10^{-6} m^2/s	50 ± 25	255	2,717	30[(a)]	107[(a)]	39[(b)]	57[(b)]	110[(a)]	278[(a)]	480[(c)]
Viscosity index	210	355	. . .	139	145			137	125	128
Pour point, °C (°F)	-53 (-63)	-66 (-87)	-15 (5)	<-60 (<-76)	-55 (-67)	-56 (-69)	-53 (-63)	<-55 (<-67)	-50 (-58)	-15 (5)
Density at 20 °C (70 °F), g/ml	1.894	1.851	1.92	. . .	. . .	0.98 (16 °C)	0.96 (16 °C)	0.85	0.8[(d)]	0.8[(d)]
Surface tension at 20 °C (70 °F), N/m	0.019	0.025	0.019	. . .	. . .	. . .	. . .	. . .	. . .	. . .
Vapor pressure, Pa										
At 20 °C (70 °F)	6.7×10^{-9}	3.9×10^{-10}	4×10^{-12}	[(e)]	. . .	[(e)]	. . .	1.4×10^{-10}[(f)]	. . .	. . .
At 100 °C (212 °F)	1.3×10^{-5}	1.3×10^{-6}	1.3×10^{-7}	. . .	. . .	. . .	. . .	. . .	. . .	. . .
At 125 °C (255 °F)	. . .	. . .	. . .	. . .	. . .	. . .	. . .	5.5×10^{-5}	. . .	. . .

(a) At 40 °C (105 °F).

(b) At 38 °C (100 °F).

(c) Extrapolated from measured viscosities at 40 and 100 °C (105 and 212 °F).

(d) Estimated based on densities of other silahydrocarbon (SiHC) samples.

(e) Note vapor pressures in Fig. 2.

(f) Extrapolated from data taken between 125 and 175 °C (255 and 345 °F)

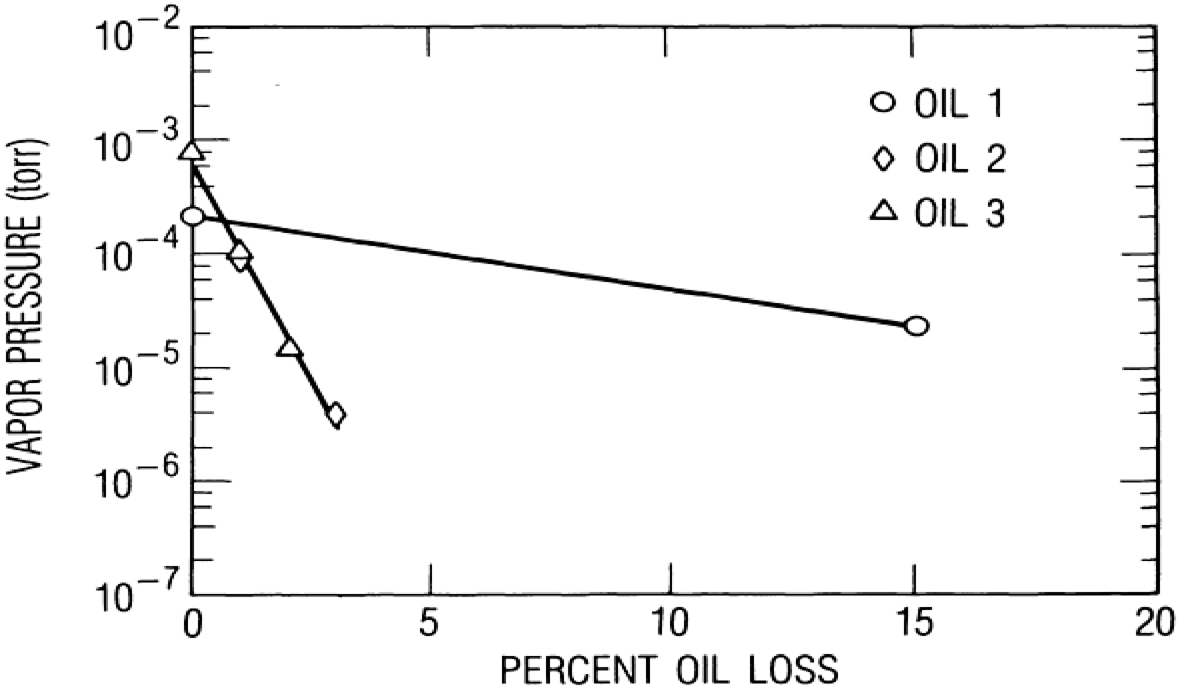


Fig. 2 Reduction in vapor pressure as a function of percent oil loss for three oils: (1) mineral oil (SRG 40); (2) poly-α-olefin (Nye 179); and (3) polyolester (Nye UC7). Source: Ref 6

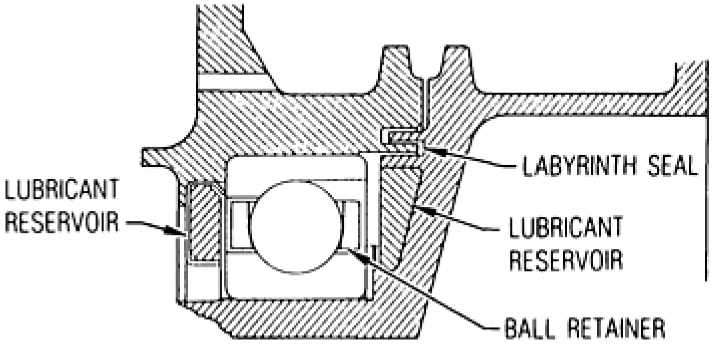


Fig. 3 Schematic of a bearing configuration showing a labyrinth seal. Source: Ref 47

The pumping speed of space vacuum is essentially infinite. There are no surfaces to reflect gas molecules back to the spacecraft. However, processes have been proposed involving photoelectric charging of emitted molecules and reattraction or collision with "ambient" molecules in the vicinity of a spacecraft and redirection toward critical surfaces (Ref 48, 49, 50). Lubricant migration by creep can be countered by anticreep barriers that are primarily made of ultralow surface energy (~0.011 N/m, or 11 dyne/cm) fluorocarbon coatings (Ref 51).

If lubricant loss does occur, either a passive or positive-feed resupply method is required. The passive method generally uses lubricant-impregnated porous solids. Oil is provided to a contact region as long as there is some positive driving force (heat or centrifugal force) to overcome the capillary forces of the porous medium. Oil-soaked phenolic retainer materials were once thought to be lubricant resupply sources for bearings. However, both theoretical and experimental studies have shown that such materials can act as sinks, further depleting the lubricant supply unless they are properly saturated with oil. This saturation process can actually take years to complete at ambient temperature (Ref 44, 52). Porous nylon-based materials, polyimides, and copolymer foams of acrylonitrile have been used as reservoirs in several mechanisms, but they can be subject to the same potential problems as phenolic retainer materials.

The positive-feed method uses positive-feed suppliers with centrifugal oilers or controlled pumps. This method has been used for higher-load (requiring larger bearing sizes) and/or longer-life mechanisms.

Several categories of liquid lubricants that either have been or could be used for vacuum/space applications include: silicone oils, mineral oils, perfluoropolyalkylethers, and other new synthetics (including poly-α-olefins, polyolesters, and multiply-alkylated cyclopentanes). Except for gyroscope applications, these lubricants generally encounter boundary contact conditions at some time during their service life. Boundary lubrication additives are available for many of these lubricants and will be reviewed after the base stocks are discussed.

Silicone Oils. The low vapor pressures and low pour points of some silicone oils led to their early use in space applications. However, these oils are only moderately effective lubricants. One problem is that some of these oils tend to form polymers on the bearing surface, which leads to torque noise. Another problem is that these oils creep readily on metal surfaces. Because of these problems and the availability of better alternatives, silicone oils would not be used on contemporary spacecraft. However, these oils are used as damper fluids and thermal conduction media in some instances.

Mineral oils that are highly refined have been a popular choice for sealed mechanisms, such as momentum wheels, reaction wheels, and despin mechanisms. Mineral oils from numerous manufacturers have been used successfully (Ref 12). A series of super-refined gyroscope lubricants is also available. These lubricants comprise a homologous group of natural polymers that allows the designer to choose a fluid having particular viscosity characteristics for a specific application (Fig. 4) (Ref 53). Mineral oils also can be formulated with antiwear and other additives.

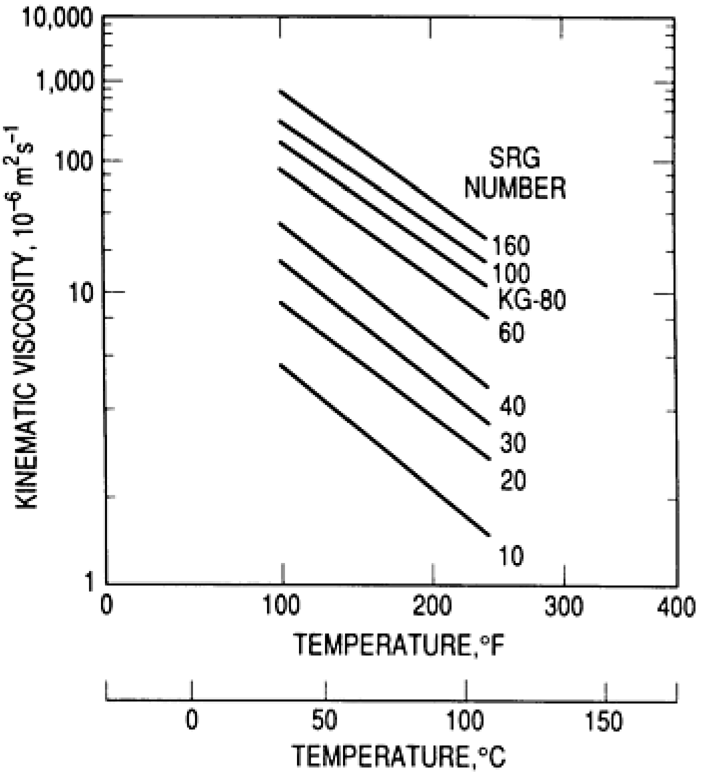


Fig. 4 Viscosity versus temperature for a homologous series of super-refined mineral oils. Source: Ref 53

Perfluoropolyalkylether (PFPE) lubricants have lower vapor pressures, lower pour points, and higher viscosity indexes than mineral oils (Table 5). They are therefore useful in space mechanisms that are not completely sealed or that are somewhat cooler (>200 K) than would be acceptable for mineral oils. In particular, one of the PFPEs has a very high viscosity index and is exceptionally useful over a wide temperature range.

PFPEs perform reasonably well under non-boundary contact conditions. However, these lubricants have definite limitations when used for applications involving boundary contact, particularly on steel surfaces (Table 6). Conventional antiwear additives do not dissolve into the PFPE fluids, although a new class of compatible additives has been reported (Ref 54). During boundary contact in the absence of additives, fluorine from the PFPE can react with iron to form FeF_3, a catalyst for the further breakdown of the polymer (Ref 55). More fluorine is released, which sustains a chain reaction.

Table 6 Factors that influence PFPE fluid degradation

Promote degradation	Retard degradation
Starved conditions	Fully flooded conditions
Low specific film thickness	High specific film thickness
Linear structure (*z*)	Branched structure (*y*)
Aluminum/titanium substrates	Hydrocarbon contamination
52100 bearing steel	440C steel and ceramic coatings
Temperatures greater than 200 °C (390 °F)	Low ambient temperatures
Sliding surfaces	Rolling surfaces
Vacuum environment	Atmospheric conditions

Lubricant degradation by polymerization leads to high bearing torque noise and excessive wear. The substrate-induced degradation can be retarded by substituting one or both of the steel surfaces with either ceramic components or ceramic-coated steel (or, presumably, by using the new additives). Both TiC- and TiN-coated steel and Si_3N_4 components have shown improved performance, as will be discussed in the section on surface modification (Ref 56).

PFPEs have extremely low surface tensions (~0.018 N/m, or 18 dyne/cm) and, therefore, creep very readily over metal and other surfaces. Because of their similar chemical structures, the lubricants also dissolve fluorocarbon coatings that are used as antimigration barriers. Commercially available PFPEs and their properties are listed in Table 5, and their molecular structures are shown in Fig. 5. The acetal groups present on the Fomblin Z25 or Braycoat 815Z polymers are particularly reactive under boundary conditions.

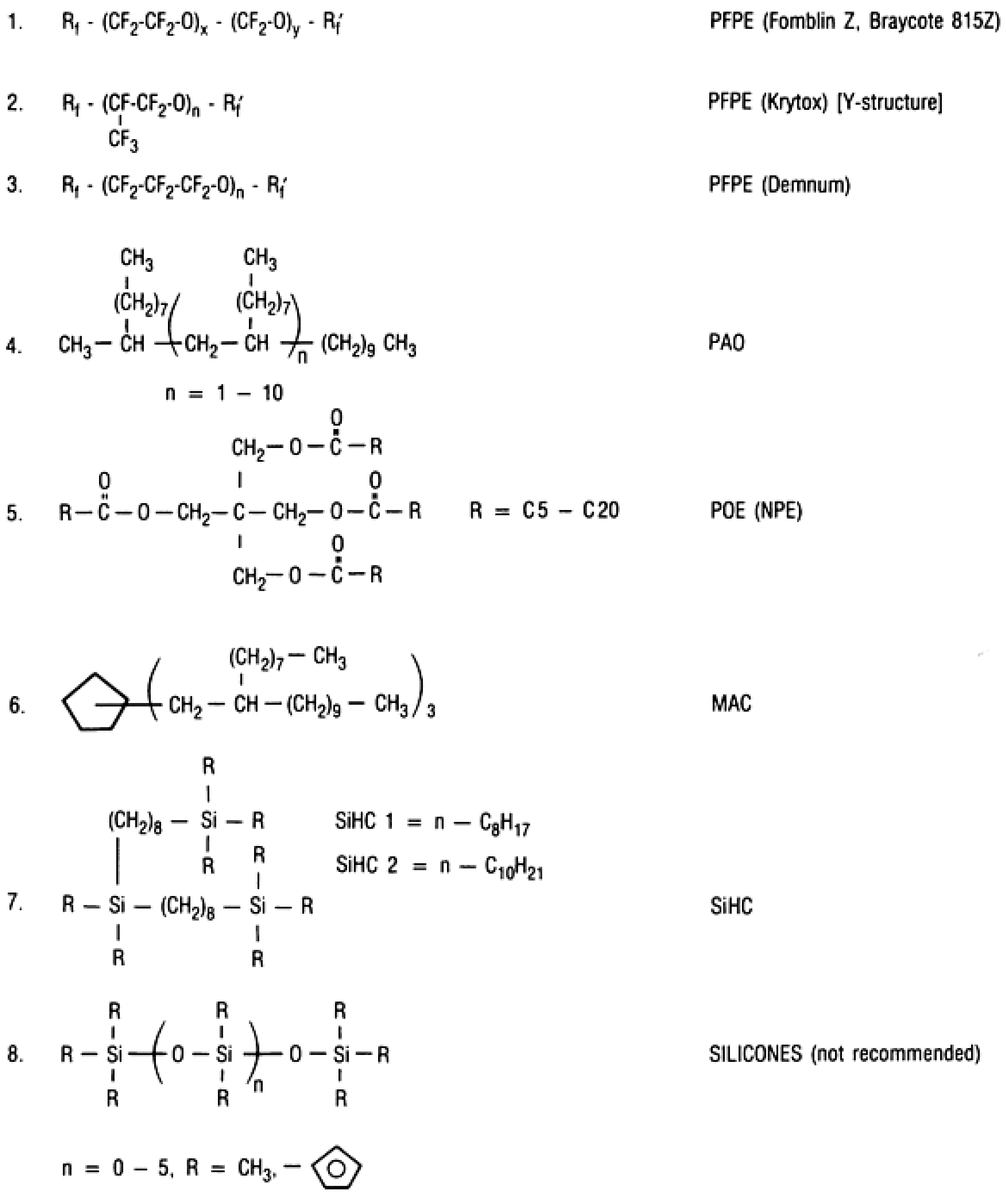


Fig. 5 Nominal molecular structure of selected fluid lubricants for space/vacuum applications

Other Synthetic Lubricants. Poly-α-olefin (PAO), polyolester (POE), multiply-alkylated cyclopentane (MAC), and other hydrocarbon polymer (HP) oils can be synthesized and blended to produce viscosity, vapor pressure, pour point, and other properties in a controlled way to suit various needs (Ref 44, 57, 58, 59, 60, 61, 62). Vapor pressures that are as low as those of linear PFPEs have not been obtained for the PAOs and POEs, but they can be lower than those of conventional mineral oils. Vapor pressure studies of the MAC oils are currently underway, and extrapolations based on measurements

at higher temperatures (125 to 175 °C, or 255 to 345 °F) suggest that room-temperature vapor pressure of at least one MAC oil should be low, as listed in Table 5 (Ref 63). Outgassing studies of selected PAOs and POEs (specifically, a neopentyl ester) show that removal of relatively high-vapor-pressure light fractions, which account for ≤3% of the as-received lubricant, reduces the vapor pressure by several order of magnitude without affecting viscosity at room temperature (Fig. 2).

These synthetic hydrocarbons can be blended with conventional additive packages to provide the same type of protection against wear, oxidation, and corrosion as achieved by natural hydrocarbons. However, for vacuum applications, the low vapor pressures of the base stocks make the additives the most volatile constituents of blended lubricants. Therefore, new additives are being developed that will be compatible with the base stocks and will have the desired low volatilities.

Laboratory screening tests have shown that synthetic hydrocarbons give the longest wear lifetimes in a simulated boundary-lubrication test facility. Bearing tests with a fixture designed to simulate the oscillatory motion of a weather scanner have shown that a PAO provides near-freezing (0 °C, or 32 °F) temperature capability and significantly outlasts both a silicone oil and a PFPE (Fig. 6) (Ref 58, 59, 60, 61, 64). PAO oils have given very good performance in lightly loaded, high-speed gyroscope bearings. Tests are presently aimed at determining the utility of these synthetic oils in more demanding applications, such as in the spin bearings of momentum and reaction wheels.

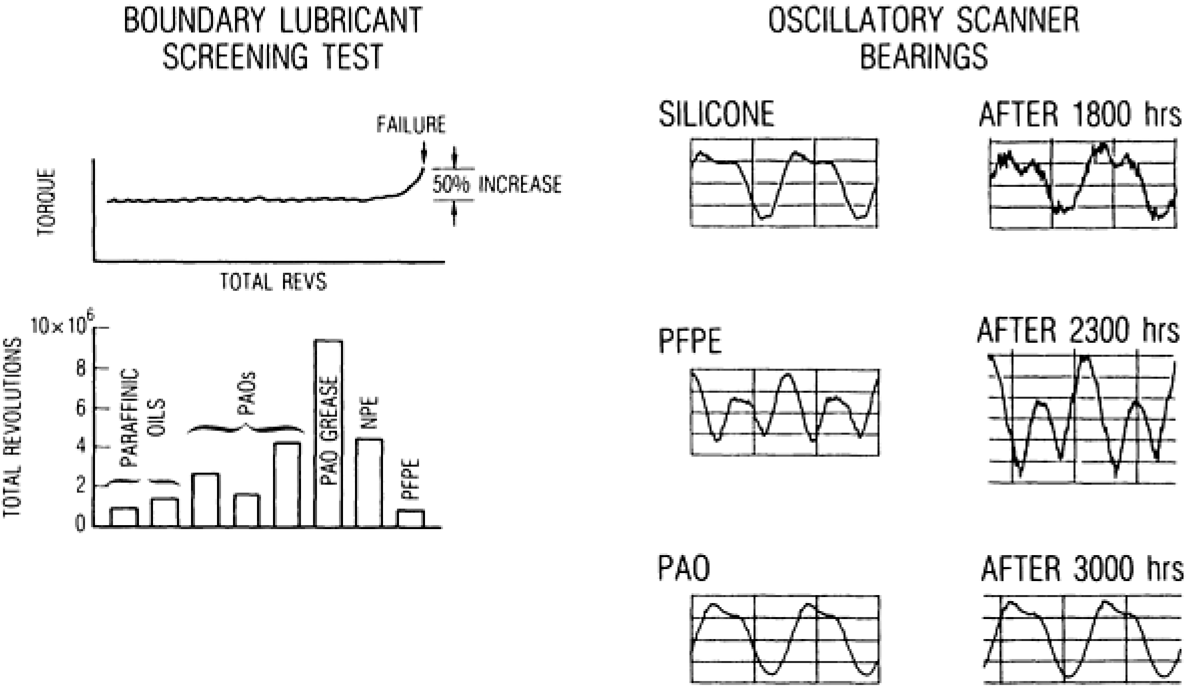


Fig. 6 Life-test results for various lubricants investigated with a boundary lubricant screening test (left) and oscillatory scanner-bearing test (right). Silicone and PFPE failed after the indicated times, whereas the PAO test continues. Source: Ref 44

Silahydrocarbons represent another relatively new class of synthetic lubricants with vapor pressures acceptable for vacuum applications and the capability to be compounded with additives (Ref 7, 8). Their tribological performance has not been thoroughly tested in specific applications, but the results of conventional four-ball and traction tests are very encouraging (Ref 65).

Additives. Liquid lubricants are formulated with additive packages in order to provide, for example, low friction and antiwear protection in elastohydrodynamic or extreme-pressure conditions, and to retard lubricant oxidation or substrate corrosion during atmospheric storage. Most of these additives were developed for use under atmospheric conditions, with oxygen and water vapor, and were compounded with base stocks on a highly empirical basis. In vacuum, once water vapor and reactive gases are removed, it is doubtful that most additives work in an identical chemical-mechanistic manner to the way they operate at atmospheric pressures. Furthermore, the nature of the interactions of these additives (which

were developed for steel substrate surfaces) with newly emerging ceramic components (such as SiC, Si_3N_4, TiC, and TiN) is unknown. Additive surface chemistry is currently an active topic of study by many tribologists and surface scientists.

Antiwear or extreme-pressure (EP) additives are often required when liquid lubricants are used in mixed or boundary regime applications in vacuum. Numerous examples exist where antiwear additives, such as tricresol phosphate, improve the operation of bearings and provide longer life. In boundary-lubrication situations, the modern designer must carefully consider the relative merits of liquid and solid lubricants. Liquid lubricants have long life and low torque, but require additives and, possibly, reservoirs and molecular vapor seals in the design. Solid lubricants operate well in boundary conditions without reservoirs or seals, but have a finite life. Another option, surface modification, is also worthy of consideration. Common EP additives include naphthenates of lead and other metals and dialkyldithiophosphates of zinc or antimony.

The role of antioxidation additives in vacuum lubricants is to prevent degradation of lubricants during storage. Common types of such additives include hindered phenols, as well as amines, such as polyphenyl α-naphthylamine.

Antirust additives can be particularly important for components stored in air that are coated with the PFPE oils, because these oils do not offer the same oxidation protection to the substrate as do other oils. Unfortunately, common antirust additives do not dissolve in the fluorocarbons any better than antiwear materials. For one of the PFPE greases, $NaNO_2$ is added by first being adsorbed onto bentonite clay and then being suspended, together with the clay, to form the grease.

Grease lubricants comprise oils compounded with a pore-network-forming thickener, such as a soap or a fine-particle suspension. Thorough descriptions of greases and their properties are provided in Ref 66 and 67. For results of an extensive testing program of greases in vacuum, see Ref 68. A very brief overview of grease lubricants is provided here for reader convenience.

Depending on the type of oil and the nature of the thickener, greases can be formulated for various applications that involve a variety of components with different types of contact (such as slow- or high-speed angular contact ball bearings, journals, and gears). Oils in greases can be from any of the categories discussed in previous sections. However, the solubility properties (chemical compatibility) of the oil will determine the selection of thickener and, hence, the grease properties. Because mineral oils and certain synthetics have good solvent properties, they can be formulated with soaps of different cations to make what are known as channeling greases. Such greases are pushed out of the way and form a path (channel) when the balls of a bearing pass through the grease. When working properly, oil will continually diffuse out of the mounds of grease on the edges of the ball path to supply lubricant to the contacting surfaces. If a grease is fluid enough that it tends to fill the spaces between balls, it is a "slumping" (non-channeling) grease. The consistency of a grease depends on the type of thickener used and the relative amounts of oil versus thickener.

Because both PAO and PFPE oils are poor solvents for soaps, greases of these oils are made by suspending fine particles of inert materials in the oils until their consistency becomes thick, like a grease. Two common thickeners of this type are a finely ground silica and another powder that is simply designated as a fluorocarbon telomer. One drawback to this type of grease is that the thickener can get into the ball path of a precision bearing and, being solid, can cause noisy operation.

The primary purpose for using grease in a vacuum application is that it can act as a reservoir for supplying oil to contacting surfaces. A bearing properly packed with grease will also suffer less oil loss that is due to either creep or physical spattering because of the physical barrier the grease can provide. However, because the lubrication properties of any grease can only be as good as those of the base oil, care must be exercised in selection of the base oil. For example, formulation of a volatile oil into a grease cannot prevent the oil from contaminating a vacuum system. Rather, a low-volatility oil must be used.

Surface Modification with and without Lubrication. The fourth approach for providing low friction and limited wear in vacuum is to use no lubricant at all. Instead, the low reactivity of the substrate surface can be relied upon to prevent cold-welding. This option can work very well for lightly loaded applications that have low-duty cycles. For mechanisms in terrestrial vacuum systems, this approach is often a tempting first choice. When a component fails to operate properly, vacuum can be broken and a lubricant can be applied and tested for the application.

Conversely, spacecraft-mechanism performance cannot be left to chance, and most mechanisms that experience more than light loads and/or have frequent use will require tribological modification in the contact zone. The modification approaches in this section do not have the same historical degree of proven success as the lubricants mentioned in the

previous sections, because the modifications are simply too new. Nonetheless, the authors believe that these emerging technologies will be used increasingly in the future, either alone or in conjunction with lubricants.

Because metals generally do not have unreactive surfaces once their passive layers are worn away, chemical modification of the surface region or complete materials substitution is required. The basic idea is to avoid metal-to-metal contact that might cause either cold-welding or adhesive wear. Ceramics often are covalently bonded materials and generally have a lower reactivity than metals (gold is one exceptionally inert metal). In particular, the carbides and nitrides of silicon and titanium (SiC, Si_3N_4, TiC, TiN) have excellent attributes: they are hard (resist plastic deformation), chemically inert (resist cold-welding/adhesive wear), and have high melting temperatures (resist chemical interdiffusion between contacting surfaces). They are also commercially available, either as ceramic components or as ceramic coatings on bearing steels.

An alternative is to ion implant C, N, or metals into steel surfaces to create a hard surface region. Another future alternative may be polycrystalline diamond coatings. However, although hard carbon films are used for some computer disk drive applications, diamond films cannot yet be applied to metals without overtempering and distortions, and their tribological properties still need extensive investigation.

Ceramic and Hard Coat Contact. TiC coatings formed by chemical vapor deposition (CVD) onto steel have been used for over a decade in the tool industry to prolong tool life. The CVD of TiC coatings is also commercially available on 440C balls for bearing applications (Ref 69). The high temperatures (>1000 °C, or 1830 °F) used in the CVD process, in which $TiCl_4$ and H_2 and N_2 gases react on the hot steel substrate to form TiC, soften the steel. Consequently, the balls have to be heat treated again to regain hardness after deposition, and anisotro phase transformations during the second heat treatment distort the balls into egg-shaped structures. The balls are then repolished to regain sphericity. Balls are available with a 9.5 mm (0.375 in.) diameter or less. Larger sizes distort to tolerances greater than the allowable coating thickness, resulting in "bald spots" after polishing. TiC-coated balls have been used without additional lubricant in the primary deployment mechanism of the recent space telescope solar array. These balls have also been used with PFPE oil for the Spacelab instrument pointing system (Ref 70).

TiN and TiC coatings produced by sputtering also have been used for years in the tool industry. The lower operating temperatures of the sputtering process, particularly some high-rate variations (Ref 71, 72) can, in principle, avoid the second heat treatment problems associated with CVD. There have been very limited studies of sputtered TiN in bearing or gear applications in vacuum. Eccentric bearing tests in vacuum showed an order of magnitude increase in life when TiN-coated components were compared to uncoated 440C steel. In both cases, the bearings were lubricated with a PFPE oil (Ref 73).

Solid ceramic parts, such Si_3N_4 balls formed by hot isostatic pressing, are commercially available with diameters up to 65 mm (2.5 in.) (Ref 74). Appropriate polishing can produce balls down to a Grade 3 finish. Grades 3, 5, and 10 are usually produced. Such balls have been used in hybrid bearings (ceramic balls and steel races) operated in ultrahigh-vacuum, either unlubricated or with solid or liquid lubricants. Eccentric bearing tests of PFPE-lubricated hybrid bearings in vacuum (Si_3N_4 balls against 440C steel races) showed an order of magnitude increase in life over an all-steel configuration. The gains observed with the Si_3N_4/steel hybrid were identical, within experimental error, to gains obtained with all TiN-coated components (Ref 63). The results emphasize the improvements that can be obtained when metal-to-metal contact is avoided, even by the elimination of only one metal surface from the contact. When used with PFPE oils, the ceramics appear to retard chemical degradation of the lubricant by iron in the steel substrate.

Although the tribological improvements of using bulk ceramics or ceramic coatings on steel appear similar, basically because metal-to-metal contact is eliminated, there are important differences in bulk properties that must be considered. The ceramic coatings are thin enough (<3 μm, or 120 μin.) that the majority of the load is carried by the steel substrate. Therefore, the modulus of the steel determines the stress levels generated. Si_3N_4 has a Young's modulus (310 GPa, or 45 × 10^6 psi) that is 50% higher than that of steel. Thus, higher contact stresses are generated in the ceramic than in the steel at any particular load. In addition, the thermal coefficient of expansion (3.5 × 10^{-6}/K) and the Poisson's ratio (0.28) must be taken into account when ceramics are combined with steel races in a given application. Tighter race conformance to the ball relative to steel bearings is sometimes required to reduce stresses, but the result can be increasing operating friction or torque.

Ion implantation of steel surfaces is a potential alternative to coatings or bulk ceramics (Ref 74, 75, 76). Energetic (100 keV) ion beams of various species (such as B, C, N, and Ti) are directed toward the substrate. Such species can form compounds in the near-surface region (<0.1 μm, or 4 μin.) or simply disrupt the surface structure (render amorphous), so that wear resistance can increase. This approach does not change the dimensions of the component. Sometimes, substrate

cooling is necessary when high-flux beams are used. The process, as originally developed, was line-of-sight, but isotropic plasma implantation techniques have become available (Ref 77). Ion implantation has been used in many different terrestrial applications. In a vacuum application, ion implantation is being tested for possible use in the main-engine bearings of the space shuttle to provide improved corrosion and wear resistance. (During use, these bearings get very hot and lubricants do not survive; stored bearings rust in condensed moisture.) A terrestrial bearing study of balls made of TiC-coated 440C, of Si_3N_4, or of ion-implanted (Ti, C) 52100 balls showed comparable improvements in performance for each modified material, compared to 52100 balls in fretting tests (Ref 78). In gyroscope spin tests, the implanted balls showed slight evidence of wear relative to the ceramic or ceramic-coated balls, but all three types performed better than the standard 52100 balls.

Specific Applications

Space Environments. Although the functions or missions of different spacecraft systems vary widely, there are many similarities in their mechanical requirements. There is a large body of literature on the design and applications experience of spacecraft mechanisms, including discussion of tribology issues (Ref 79). Table 7 lists mechanical components often found on spacecraft and is arranged according to the type of mechanical contact. The list is not meant to be exhaustive, but is typical of components that have exhibited specific problems (Ref 80). Recent articles (Ref 81) are all devoted to various aspects of tribology of space systems throughout the world. The space program of each country is treated individually, and there are some general articles on different technologies.

Table 7 Tribology components: requirements and technology "solutions" for space applications

Mechanical components	Requirements									Technology "solutions"					
	Low friction	**Low wear rate**	**Extreme environment**	**Electrical/thermal conductivity**	**Periodic motion**	**Gas/vacuum compatibility**	**Storage**	**Noncontaminating**	**Controlled backlash**	**Fluid lubes**	**Solid lubes**	**Hard coats**	**Composite materials**	**Additives**	**Ceramic bulk materials**
Sliding contact															
Clamps/latches	A	. . .	A	. . .	. . .	A	A	A	A	. . .	D	. . .	. . .	. . .	. . .
Slip rings	A	B	B	A	. . .	. . .	A	. . .	. . .	C	D	. . .	. . .	. . .	. . .
Potentiometers	A	A	A	A	A	. . .	. . .	. . .	A	. . .	D	. . .	D	. . .	. . .
Seals	A	A	A	. . .	A	A	. . .	A	. . .	. . .	C	. . .	D	. . .	. . .
Bearing retainers	A	A	. . .	. . .	A	. . .	A	. . .	. . .	D	C	. . .	D	. . .	. . .
Telescoping joints	A	B	A	A	A	. . .	A	A	. . .	. . .	D	D	D	. . .	D
Gas/magnetidc bearings	A	. . .	A	. . .	A	A	. . .	. . .	. . .	. . .	D	D	D	. . .	. . .
Rolling contact															
Spin bearings	A	A	. . .	A	. . .	. . .	A	. . .	. . .	D	. . .	D	D	. . .	D
Gimbal bearings	A	B	A	. . .	A	. . .	A	A	. . .	C	D	D	D	C	D
Sensor support bearings	A	A	A	A	A	. . .	A	A	. . .	D	D	D	D	D	D
Motor, synchronous	A	A	A	A	. . .	. . .	A	A	A	C	D	. . .	. . .	. . .	. . .
Solar array bearings	A	. . .	A	A	A	. . .	A	A	. . .	. . .	D	C	D	. . .	C
Mixed regime															
Gears	A	A	A	A	. . .	A	. . .	A	A	D	D	D	. . .	D	C
Harmonic drives	A	A	A	. . .	. . .	. . .	. . .	. . .	A	D	C	D	. . .	D	C

Legend: A, critical requirement; B, less stringent requirements; C, technology that would probably improve performance; D, technology that would improve performance

Source: Ref 44

It is important for the reader to understand that the field of space tribology is undergoing major advances in lubricant technology. Such advances are being driven by two trends: (1) the required lifetimes of spacecraft are increasing, making some of the past lubricant practices inadequate; and (2) mechanical component failures are starting to become the life-limiting systems on spacecraft, because the traditional failure points (power systems, electronics, and contamination problems) are using newer technologies that exceed the lifetimes of the tribosystems (Ref 45, 46, 48, 49, 50). The tribology advances include new or improved tribomaterials and the generation of more extensive test data to qualify these tribomaterials for various applications. Table 7 also lists materials technologies either in use or available for implementation on various spacecraft components. The different symbols in the table correspond to a critical or less stringent requirement or to a preferred or alternative technology. The reader is referred to the references for more details concerning vacuum applications.

Terrestrial ultrahigh vacuum (UHV) environments, or chambers, are used primarily by:

- Surface scientists conducting chemical experiments or analytical measurements (such as spectroscopy, diffraction, or microscopy), or both, on the first few atomic surface layers of materials
- Scientists and engineers synthesizing or fabricating materials with features or structures of micrometer to nanometer scale
- Engineers testing hardware intended for vacuum applications

All three groups require extremely clean, stable environments to obtain reproducible results. Contamination from lubricants in the form of vapors, migrating liquid lubricant molecules, or particulate debris is not acceptable. Designers try, as much as possible, to keep mechanical components of such systems external to the vacuum and rely on feed-throughs, magnetic couplings, or bellows to transmit or facilitate component motion. However, UHV systems often contain manipulators to move materials within the chamber or between chambers. Rotatable fixtures are also required in line-of-sight deposition processes to adequately expose nonplanar substrates to the deposition flux. If the manipulators are used frequently, undergo moderate to heavy stresses, or have tight-tolerance components, some lubrication is generally required to avoid frequent mechanism repair (which necessitates breaking vacuum). Lubricants for pumps are not covered in this article, because that technology is well developed, and the pump manufacturers are quite capable and flexible in meeting customer requirements for specific applications.

The general strategy that has been used is to avoid metal-to-metal contact and to have minimum lubrication. For example, PTFE and polyimide polymers (pure or composites) have been used to make bearing bushings for UHV manipulators. Recently, Si_3N_4 balls have been used to make hybrid bearings for UHV mechanisms. (Presumably, TiC- or TiN-coated materials would also work.) The Si_3N_4 has been used without lubricant or with spray-deposited WS_2 for more demanding applications (Ref 82). WS_2 has also been used on steel. Sputter-deposited MoS_2 has been used on deposition fixtures and on a cryogenic (helium) manipulator for an acoustic microscope (made of steel), resulting in lower operating torque of these mechanisms.

Grease (polytetrafluoroethylene, or PTFE, base) has also been used in a robot manipulator operated in a molecular beam epitaxy (MBE) system (Ref 83). The MBE robot has many gears, bearings, and lead screws. Over 20 robots have been built, and some have operated over 5 years without maintenance. The grease was selected instead of solid lubricants partly because of a desire to avoid particle generation. Micrometer-sized particles have unusual migration tendencies in vacuum (easily propelled by electrostatic charges) and are devastating to the fabrication of submicrometer devices. The particle-generation properties of solid lubricants as a function of film properties (for example, porosity) have not been systematically studied. Further work is needed and is in progress (Ref 84). The designer should place mechanisms below critical surfaces in vacuum, so that gravity will move away debris from the sites of inevitable wear. Containment barriers should also be considered that are interposed in the line of sight between critical surfaces and the lubricated sections of mechanisms.

PTFE-based greases are also used to lubricate Viton seals between UHV and air on sliding manipulators. Such seals are usually in multiple series, with differential pumping between each seal.

When selecting tribomaterials, the designer should review their temperatures stabilities, load-bearing capabilities, moisture sensitivities, and achievable tolerances. Obviously, the vapor pressure of the tribomaterial must be below desired system operating pressure. Because UHV systems are routinely baked out, the vapor pressure and load-carrying capacities of the tribomaterials must not degrade by such heating. Moisture sensitivity is important if the mechanism is routinely stored and/or operated in air. Polyimides are extremely sensitive to water vapor (these polymers absorb as much as 2%

moisture, and minutes or hours of moisture exposure will require subsequent bake-out), whereas PTFE is not. However, polyimides have load-bearing capabilities superior to those of PTFE. MoS_2 is more sensitive to water vapor exposure than WS_2 (MoS_2 will significantly oxidize over a period of months to a year), yet MoS_2 has better sliding wear endurance in vacuum. (MoS_2 oxidation can be effectively avoided at atmospheric pressure by storage of the lubricated component in a dry or inert gas, desiccated environment.) Sputter deposition can yield solid-lubricant coatings to better tolerance than either burnishing or spray methods. Sputter-deposited films can also be prepared that will resist oxidation upon standing (not operating) in air for many months, and such films should be considered for conventional uses. Bonded films, though subject to greater debris formation, are generally much more resistant to oxidation. A typical procedure for applying bonded films is to run the film in (burnish) after the normal application and then remove any debris (with a vacuum brush) before using.

References

1. H.M. Briscoe, *Tribol. Int.,* Vol 23 (No. 2), 1990, p 67
2. D.R. Peplinski, G.S. Arnold, and E.N. Borson, *Introduction to: Simulation of Upper Oxygen Satellite Exposure to Atomic Oxygen in Low Earth Orbit,* No. 2340, 13th Space Simulation Conference, National Space and Aeronautics Association, Oct 1984, p 133-145
3. H.E. Sliney, *J. Vac. Sci. Technol. A.,* Vol 4, 1986, p 2629
4. C. DellaCorte and H.E. Sliney, *ASLE Trans.,* Vol 30, 1987, p 77
5. I.L. Singer, *Mater. Res. Soc. Symp. Proc.,* Vol 140, 1989, p 215
6. D.J. Carré, "A Model to Describe Oil Loss Under Vacuum Conditions," TR-0090 (5945-03), The Aerospace Corporation, in press
7. C. Taborski and C.E. Snyder, Jr., "Silahydrocarbons: I. Synthesis and Properties of a Class of Thermally Stable Fluids," AFWAL-TR-80-4072, Wright Aeronautical Laboratories, Nov 1980
8. S.K. Sharma, C.E. Snyder, Jr., and L.J. Gschwender, *Ind. Eng. Chem. Res.,* Vol 29, 1990, p 1858
9. K. Miyoshi and D. H. Buckley, *ASLE Trans.,* Vol 27 (No. 1), 1983, p 15
10. F.P. Bowden and D. Tabor, *The Friction and Lubrication of Solids,* Clarendon Press, Oxford, 1958
11. E.W. Roberts, *Thin Solid Films,* Vol 181, 1989, p 461
12. E.L. McMurtrey, "Lubrication Handbook for the Space Industry. Part A: Solid Lubricants; Part B: Liquid Lubricants," TM-86556, NASA George C. Marshall Space Flight Center, 1985
13. R.A. Rowntree and M.J. Todd, *Mater. Res. Soc. Symp. Proc.,* Vol 140, 1989, p 21
14. *Guiness Book of World Records,* Bantam, 1988, p 182
15. M.N. Gardos, *Tribol. Trans.,* Vol 31 (No. 2), 1988, p 214
16. E.W. Roberts and W.B. Price, *Mater. Res. Soc. Symp. Proc.,* Vol 140, 1989, p 251
17. E.W. Roberts, *Tribol. Int.,* Vol 23 (No. 2), 1990, p 95
18. W.E. Jamison, *Proceedings of the Third International Conference on Solid Lubrication,* ASLE SP-14, Society of Tribologists and Lubrication Engineers, 1984, p 73
19. T. Spalvins, *ASLE Trans.,* Vol 14, 1971, p 267
20. T. Spalvins, *ASLE Trans.,* Vol 17, 1973, p 1
21. L.E. Pope, L.L. Fehrenbacher, W.O. Winer, Ed., New Materials Approaches to Tribology: Theory and Applications, *Mater. Res. Soc. Symp. Proc.,* Vol 140, 1989, p 215-284
22. P.D. Fleischauer, *Thin Solid Films,* Vol 154, 1987, p 309
23. J.-Ph. Nabot, A. Aubert, R. Gillet, and Ph. Renaux, *Surf. Coat. Technol.,* Vol 43/44, 1990, p 629
24. C. Muller, C. Menoud, M. Maillat, and E.E. Hintermann, *Surf. Coat. Technol.,* Vol 36 (No. 1-2), 1988, p 351
25. L.E. Pope and J.K.G. Panitz, *Surf. Coat. Technol.,* Vol 36 (No. 1-2), 1988, p 341
26. M.R. Hilton, R. Bauer, and P.D. Fleischauer, *Thin Solid Films,* Vol 188, 1990, p 219
27. V. Buck, *Thin Solid Films,* Vol 139, 1986, p 157
28. M.R. Hilton and P.D. Fleischauer, *Mater. Res. Soc. Symp. Proc.,* Vol 140, 1989, p 227

29. H. Kuwano and K. Nagai, *J. Vac. Sci. Technol. A,* Vol 4, 1986, p 2993
30. R.N. Bolster, I.L. Singer, J.C. Wegand, S. Fayeulle, and C. R. Gossett, Preparation, Analysis, and Tribological Behavior of IBAD MoS_2 Films, *Surf. Coat. Technol.,* Vol 46, 1991, p 207-216
31. B.C. Stupp, *Thin Solid Films,* Vol 84, 1981, p 257
32. B.C. Stupp, in *Proceedings of the Third International Conference on Solid Lubrication,* ASLE SP-14, Society of Tribologists and Lubrication Engineers, 1984, p 217-222
33. M.R. Hilton, R. Bauer, S.V. Didziulis, M.T. Dugger, J. Keem, and J. Scholhamer, Structural and Tribological Studies of MoS_2 Solid Lubricant Films Having Tailored Metal-Multilayer Nanostructures, *Surf. Coat. Technol.,* Nov 1990
34. P.D. Fleischauer and R. Bauer, *Tribol. Trans.,* Vol 31 (No. 2), 1988, p 239
35. F.P. Bowden and D. Tabor, *The Friction and Lubrication of Solids,* Part II, Clarendon Press, Oxford, 1958, p 191
36. I. Iliuc, *Tribology of Thin Layers,* Elsevier Science Publishers B.V., 1980, p 64
37. A.A. Conte, *ASLE Trans.,* Vol 26 (No. 2), 1982, p 200
38. M.N. Gardos, *Tribol. Trans.,* Vol 31 (No. 2), 1988, p 248-249
39. M.N. Gardos and C.R. Meeks, "Solid Lubricated Rolling Element Bearings, Part II: Turbine Bearings and the Associated Solid Lubricants Research; Volume 1: Summary," AFWAL-TR-83-4129, Hughes Aircraft Company, 1984, p 49-60 and 204-216
40. M.N. Gardos, Self-Lubricating Composites for Extreme Lubricating Conditions, *Friction and Wear of Polymer Composites,* F. Klaus, Ed., Elsevier Science Publishers B.V., 1986, p 397-447
41. R.L. Fusaro, *Tribol. Int.,* Vol 23 (No. 2), 1990, p 105
42. R.L. Fusaro, *Tribol. Trans.,* Vol 31 (No. 2), 1988, p 174
43. P. Sutor and M.N. Gardos, *Proceedings of the Third International Conference on Solid Lubrication,* ASLE SP-14, STLE, 1984, p 258
44. P.D. Fleischauer and M.R. Hilton, *Mater. Res. Soc. Symp. Proc.,* Vol 140, 1989, p 9
45. E.W. Roberts, "A Review of Sliding Electrical Contacts for Space Applications," ESA report under Contract ESA 4099/79/NL/PP, European Space Agency/European Space Tribology Laboratory, 1981, p 52
46. E.W. Roberts, Sliding Electric Contacts in Space: Observations on Existing Technology and New Trends in Low-Speed Applications, *Proceedings of the First European Space Mechanisms and Tribology Symposium,* ESA SP-196, Neuchatel, 1983, p 3-10
47. M.N. Gardos, *ASLE Trans.,* Vol 17 (No. 4), 1974, p 237
48. D.F. Hall and J.N. Wakimoto, in *AIAA 19th Thermoplastics Conference,* AIAA-84-1703, American Institute of Aeronautics and Astronautics, June 1984
49. J.J. Scialdone and A.E. Hedin, *J. Geophys. Res.,* Vol 83, 1978, Al 195
50. D.P. Cauffman, "Ionization and Attraction of Neutral Molecules to a Charged Spacecraft," TR-0081 (6960-06)-2, The Aerospace Corporation, 1980
51. E.V. Zaretsky, *Tribol. Int.,* Vol 23 (No. 2), 1990, p 75
52. L. Dormant and S. Feuerstein, *J. Spacecr. Rockets,* Vol 13, 1976, p 306, 755
53. J.W. Kannel and K.F. Dufrane, Rolling Element Bearings in Space, *Twentieth Aerospace Mechanisms Symposium,* NASA CP-2423, 1986, p 125-132
54. S.K. Sharma, L.J. Gschwender, and C.E. Snyder, *J. Syn. Lubr.,* Vol 7 (No. 1), 1990, p 15
55. D.J. Carré, *Tribol. Trans.,* Vol 31 (No. 4), 1988, p 437
56. D.J. Carré, *Surf. Coat. Technol.,* Vol 43-44, 1990, p 609
57. A. Beerbower, What's So Hot About Formulated Synthetics?, *Tribology in the 80's,* Vol 1, NASA Conference Publication 2300, 1984, p 477-499
58. "Mobil SHF Base Stocks," Technical Bulletin SHFBSD72, Mobil Chemical Company
59. "Synthetic Hydrocarbon Precision Lubricating Oils," technical bulletin of Nye Specialty Lubricants

60. "NPT 4 Lubricating Oil, Synthetic for Low and High Temperatures," technical bulletin of Bray Oil Company
61. "Pennzane Synthesized Hydrocarbon Fluid X 2000, Technical Bulletin CF-BFI, Pennzoil Products Company, 1991"
62. C.L. Venier and E.W. Casserly, *Lubr. Eng.,* Vol 47 (No. 7), 1991, p 586-591
63. C.L. Venier, Pennzoil Products Company, private communication, May 1991
64. D.J. Carré, P.D. Fleischauer, C.G. Kalogeras, and H.D. Marten, *J. Tribology,* Vol 113, 1991, p 308
65. S.K. Sharma, C.E. Snyder, Jr., and L.J. Gschwender, Tribological Properties of Some Advanced Space Lubricants, *Tribol. Trans.,* in press
66. C.J. Boner, *Modern Lubricating Greases,* Scientific Publications Ltd., Great Britain, 1976
67. E.R. Booser, Ed., *CRC Handbook of Lubrication (Theory and Practice of Tribology) Volume II: Theory and Design,* CRC Press, 1984
68. E.L McMurtrey, *High Performance Liquid and Solid Lubricants--An Industrial Guide,* Noyes Data Corporation, 1987
69. H.J. Boving and H.E. Hintermann, *Thin Solid Films,* Vol 153, 1987, p 253
70. H.J. Boving and H.E. Hintermann, *Tribol. Int.,* Vol 23 (No. 2), 1990, p 129
71. W.D. Sproul, *Thin Solid Films,* Vol 107, 1983, p 141
72. T.P. Chang, H.S. Cheng, W.A. Chiou, and W.D. Sproul, A Comparison of Fatigue Failure Morphology Between TiN Coated and Uncoated Lubricated Rollers, *Tribol. Trans.,* Vol 34 (No. 3), 1991, p 408-416
73. J.W. Lucek and J.G. Hannoosh, Field Experience in Ceramic Bearings, *Engineered Materials for Advanced Friction and Wear Applications,* ASM International, 1988, p 205-208
74. F.A. Smidt, B.D. Sartwell, and S.N. Bunker, *Mater. Sci. Eng.,* Vol 90, 1990, p 385
75. L.E. Pope, J.A. Knapp, and D.M. Follstaedt, *Surf. Coat. Technol.,* Vol 36, 1988, p 361
76. C.A. Carosella, I.L. Singer, R.C. Bowers, and G.R. Gossett, *Ion Implantation Metallurgy,* C.M. Preece and J.K. Hirvonen, Ed., AIME, 1980, p 103
77. J.R. Conrad, S. Baumann, R. Flemming, and G.P. Meeker, *J. Appl. Phys.,* Vol 65 (No. 4), 1989, p 1707
78. R.A. Hanson, The Effect of Advanced Materials Technologies on Wear in Instrument Bearings, *Ball Bearing Symposium Proceedings,* DoD/Instrument Bearing Working Group and C.S. Draper Laboratories, 1987
79. *Proceedings of the Aerospace Mechanisms Symposium,* National Aeronautics and Space Administration, annual issues
80. P.D. Fleischauer and M.R. Hilton, *Tribol. Int.,* Vol 23 (No. 2), 1990, p 135
81. H.M. Briscoe, Ed., *Tribol. Int.,* Vol 23 (No. 2), 1990, p 65-160
82. J.W. Lucek, Cerbec Corporation, private communication, Sept 1990
83. J. Ayers, Texas Instruments, private communication, Nov 1990
84. L.E. Pope and M.T. Dugger, Sandia National Laboratories, private communication, Jan 1991

Internal Combustion Engine Lubricants

Syed Q.A. Rizvi, The Lubrizol Corporation

Introduction

AN INTERNAL COMBUSTION engine burns fuel within itself. The products of combustion drive the mechanism that converts the latent energy of the fuel into mechanical energy.

Most engines currently used in transportation and industry are of the internal combustion type. The prime reasons for their popularity are their high thermal efficiency and their light weight, in relation to their power capability.

There are four general types of internal combustion engines: four-stroke cycle engines, two-stroke cycle engines, rotary engines, and continuous combustion gas turbine engines (Ref 1). The four-stroke and two-stoke cycle engines are based on the reciprocating piston principle, and are the most common. They include both compression ignition (CI) and spark ignition (SI) types. Although the CI and SI engines have many common features, they differ in the manner in which the fuel and the air are mixed. In CI engines, the intake system supplies only air, which is compressed before the fuel is injected into the combustion chamber. In SI engines, the fuel and the air are mixed in either the carburetor or intake port (fuel injection) prior to entry into the combustion chamber.

Four-stroke CI and SI engines are used in most automotive applications. Depending on the cylinder arrangement, four-stroke engines can be subcategorized as in-line, V (vee), opposed-piston, and radial types. Two-stroke SI engines are commonly used in lightweight applications (outboard motors, motor scooters, motor cycles, snowmobiles, and chain saws), whereas two-stroke CI engines are used in on-highway trucks, city buses, low-speed marine applications, and some railroad diesel applications.

Lubricants and Lubrication

All engines must be lubricated in order to preserve the integrity of the system for its designated lifetime. The extreme temperatures in internal combustion engines make lubrication complex. The lubricants that are suitable for use in these engines are expected to reduce friction, dissipate heat from internal parts, minimize deposit formation, and prevent corrosion and wear (Ref 2).

The world lubricant market for 1990 was estimated at 10.2 billion gallons, of which North American consumption represented over 28% of the total (valued at $6.5 billion). The centralized economies consumed over 29%; Western Europe, over 16%; Japan and Australia, over 14%; Latin America, nearly 7%; and Africa and the Middle East, over 4%. Engine oils, that is, internal combustion engine lubricants, accounted for 51% of the total lubricant use; general machine fluids accounted for 23%; process oils, 17%; metalworking fluids, 7%; and greases, 2%.

In most applications, the lubricant or oil is delivered from a reservoir to those parts of the equipment that require lubrication and cooling. The lubricating system can be full-pressure, splash, or modified-splash types (Ref 3, 4). Most automotive engines use a full-pressure system, in which the oil is pumped from the oil sump to the main bearing and connecting rods, and then up the connecting rods to the piston pin. In overhead valve engines, a portion of the pumped oil travels through push rods (in some cases), over rocker arms, past valve stems, and down the valve guides. In many engines, the cylinder walls and piston pins depend on splash lubrication by the oil that is thrown off the main bearing (Ref 4). Figure 1 depicts the lubrication system of a cam-in-block overhead valve engine.

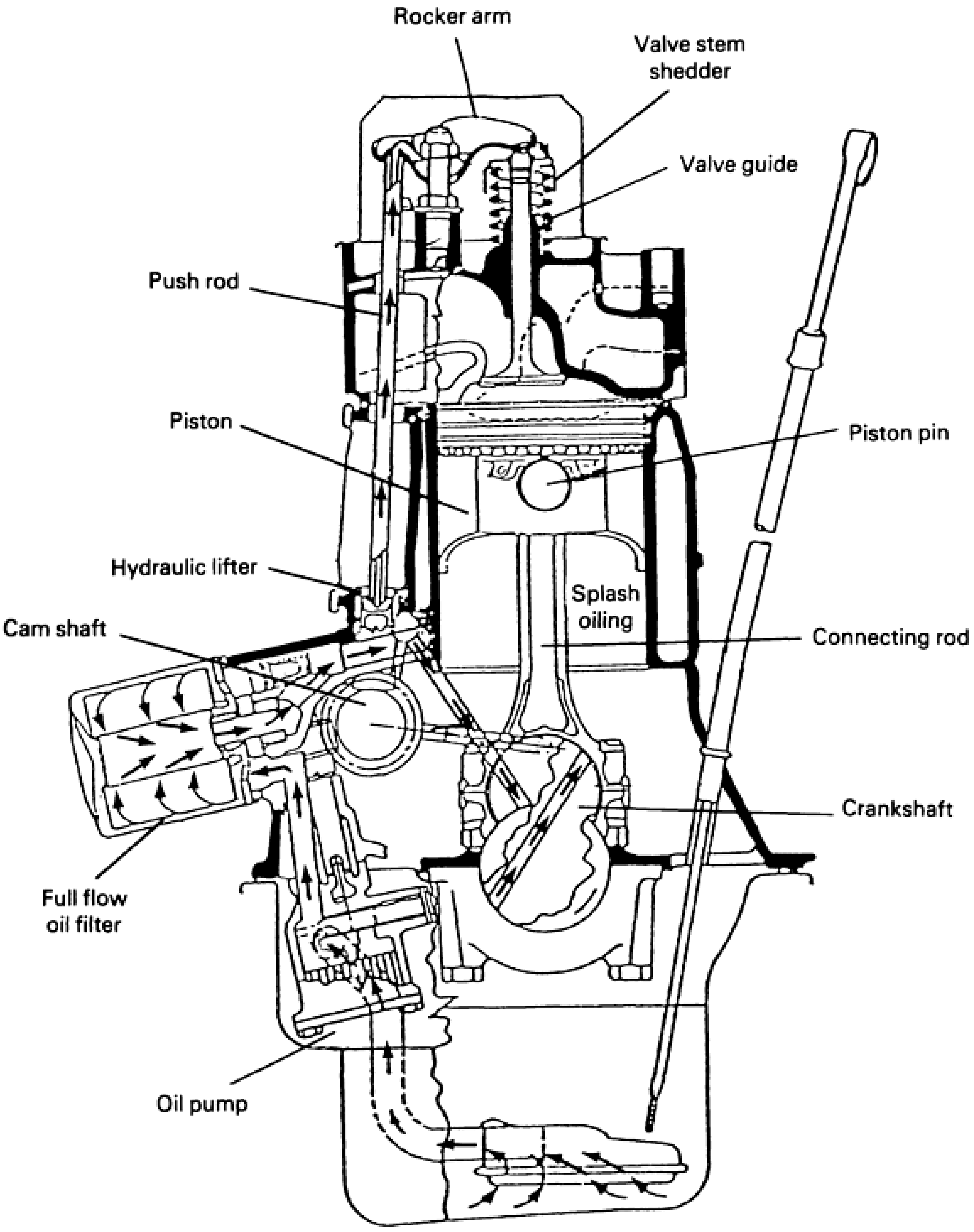


Fig. 1 Lubrication mechanisms in cam-in-block overhead block valve engine. Source: Ref 4

The main purpose of lubrication is to reduce friction by forming a film between two moving surfaces. The strength and durability of this film is related to the viscosity of the lubricant and to the speed and load experienced by the moving surfaces.

The relationship of coefficient of friction (μ) and oil film thickness with lubricant viscosity (*Z*), equipment speed (*N*), and equipment load or pressure (*P*) are shown as the Streibeck curve in Fig. 2 (Ref 5). The ratio of $(ZN)/P$ is related directly to the oil film thickness, but inversely to the coefficient of friction. This implies that high lubricant viscosity, high equipment speed, and low equipment load will allow the formation of a thick lubricant film and, hence, the equipment

will encounter little or no friction. (The initial drop in the coefficient of friction while moving from fluid-film to mixed-film lubrication reflects a decrease in viscous drag because of a decrease in lubricant viscosity.) Conversely, low lubricant viscosity, low equipment speed, and high equipment load will create a situation where the film thickness will be inappropriate and the equipment will encounter high friction.

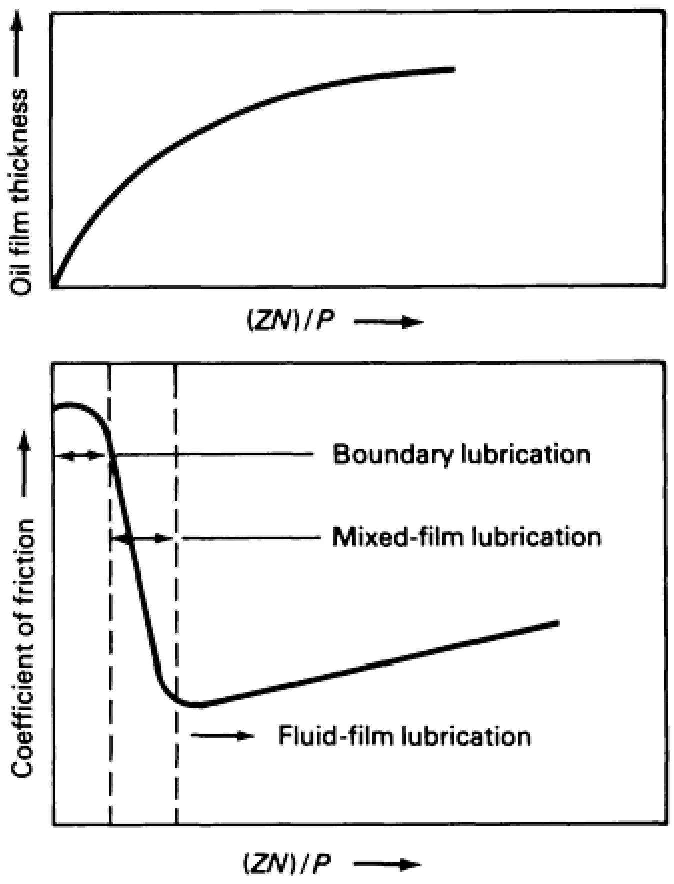


Fig. 2 Streibeck curve showing lubrication regimes. Source: Ref 5

Depending on the lubricating environment, lubrication can be divided into fluid-film, boundary film, and mixed-film types. The three states of lubrication are shown in Fig. 3 (Ref 6).

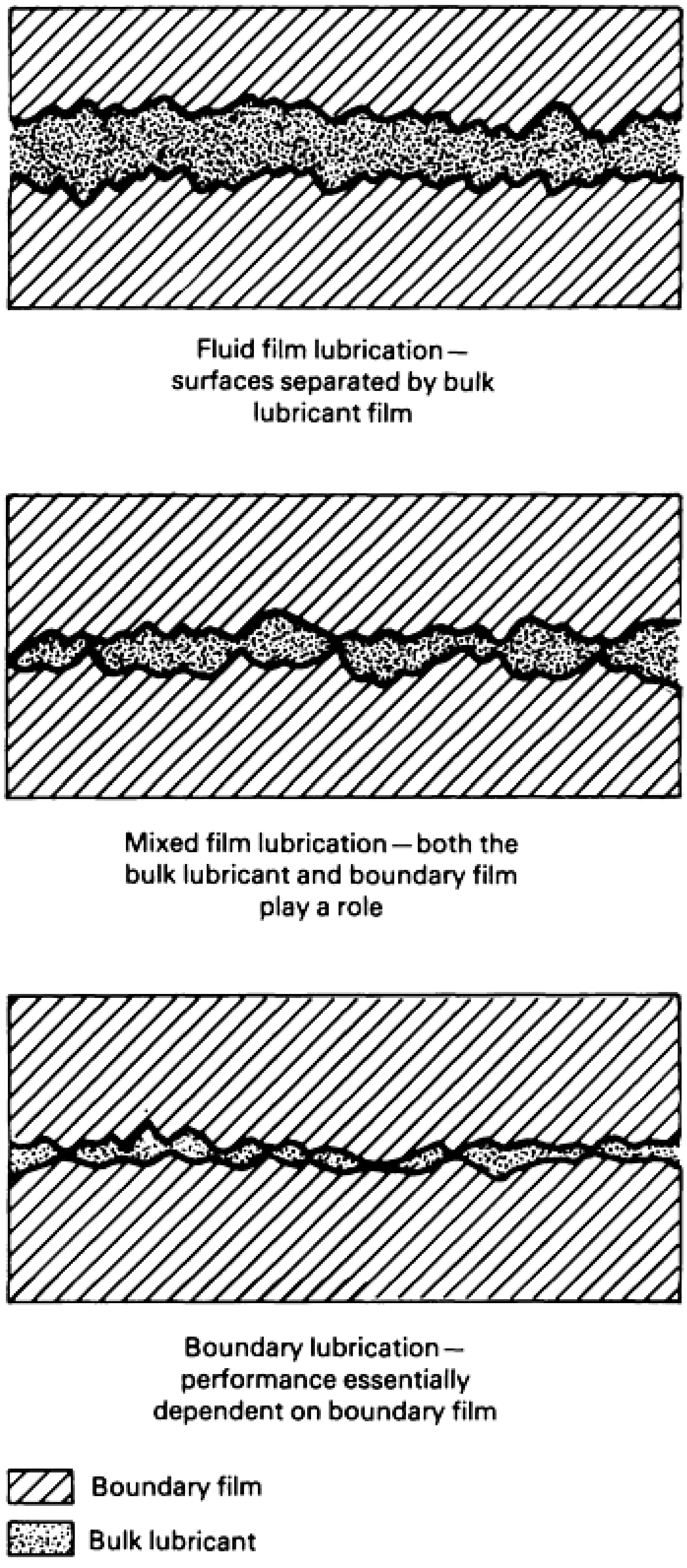


Fig. 3 Lubricant film thickness versus lubrication regimes. Source: Ref 6

Fluid-film lubrication, also known as hydrodynamic lubrication, is most desirable. This type of lubrication depends on the viscosity of the lubricant, and is effective only when the load in the contact zone is low. Under these circumstances, the sliding surfaces are separated by a lubricant film that is several times the thickness of the surface roughness.

Another type of hydrodynamic lubrication, referred to as elastohydrodynamic lubrication (EHD) (Ref 7, 8), commonly occurs during automotive gear operations. The lubricant is exposed to high contact pressures and undergoes a large viscosity increase. This results in an extremely rigid lubricant film, which causes elastic deformation of the surfaces in the lubricating zone.

Boundary lubrication simulates the other extreme of the lubricating environment. High loads and very slow speeds produce extreme pressures that can lead to the lack of effective lubrication and, hence, promote maximum metal-to-metal contact. If not controlled, the resulting dry metallic friction will cause catastrophic wear and, ultimately, total seizure. Reactive chemicals, called antiwear and extreme pressure agents, provide protection in this kind of lubricating environment. Although boundary lubrication is encountered in certain parts of the engine, such as the valve train, cylinder bores, and piston rings, most of the lubrication in an engine is hydrodynamic in nature.

Mixed-film lubrication falls between the two extremes described above and contains characteristics of each. There are regions of no metal-to-metal contact, as well as regions of metal-to-metal contact.

Specifications

The physical characteristics and performance of engine oils are defined by the Society of Automotive Engineers (SAE) viscosity grades, American Petroleum Institute (API) service classifications, U.S. military specifications, and the original equipment manufacturer (OEM) requirements.

Viscosity is one of the most important properties of a lubricant. It can be defined as the resistance of a lubricant to flow. Viscosity classification systems for engine oils are described by SAE Standards J 300 (Ref 9) and J 1536 (Ref 10). The basic viscosity grade categories for engine oils are determined by the Crankcase Classification System devised by SAE. The SAE uses test methods that are approved by the American Society for Testing and Materials (ASTM). There are a total of 11 viscosity grades, ranging from 0W to 25W, and from 20 to 60. Each viscosity grade must meet a number or requirements.

Monograde viscosity oils with the letter "W" (winter) must meet the minimum borderline pumping temperature, maximum low-temperature viscosity (at a prescribed temperature), and the minimum kinematic viscosity requirements at 100 °C (212 °F). ASTM test procedures D 2802 (modified) and D 445 are used to determine low- and high-temperature viscosities, and ASTM D 3892 is used to determine the borderline pumping temperature. Monograde oils without the letter "W" must meet the high-temperature (100 °C) viscosity requirement only. Multigrade oils should satisfy the appropriate requirements of both the "W" and the non-"W" grades. The winter requirement of being pump-able at low temperatures and the nonwinter requirement of sufficient viscosity at high temperatures ensure proper lubrication and protection of engine parts during all-season operation. These requirements, along with those of the U.S. military, are summarized in Table 1 (Ref 11).

Table 1 Viscosity requirements for engine oils

Property	SAE viscosity grades for crankcase oils, SAE J 300										
	0W	**5W**	**10W**	**15W**	**20W**	**25W**	**20**	**30**	**40**	**50**	**60**
Viscosity[a], max, Pa · s	3.25[b]	3.5[c]	3.5[d]	3.5[e]	4.5[f]	6.0[g]	. . .	. . .	. . .	. . .	. . .
Borderline pumping temperature[h], max, °C (°F)	-35 (-30)	-30 (-20)	-25 (-15)	-20 (-5)	-15 (5)	-10 (15)	. . .	. . .	. . .	. . .	. . .
Viscosity[i] at 100 °C (212 °F), mm²/s											
Min	3.8	3.8	4.1	5.6	5.6	9.3	5.6	9.3	12.5	16.3	21.9
Max	. . .	. . .	. . .	. . .	. . .	. . .	<9.3	<12.5	<16.3	<21.9	<26.1
Stable pour point, max, °C (°F)	. . .	-35 (-30)	-30 (-20)	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .

Property	Military grades					
	10W	**30**	**40**	**5W-30**	**10W-30**	**15W-40**
Viscosity at 100 °C (212 °F), mm²/s						
Min	5.6	9.3	12.5	9.3	9.3	12.5
Max	<7.4	<12.5	<16.3	<12.5	<12.5	<16.3
Cold cranking simulator (CCS) viscosity, Pa · s						
Min	3.5[j]	. . .	. . .	3.25[b]	3.5[j]	3.5[c]

Max	3.5[c]	. . .	. . .	3.5[j]	3.5[c]	3.5[e]
Mini-rotary viscometer (MRV) borderline pumping temperature, max, °C (°F)	-25 (-15)	. . .	. . .	-30 (-20)	-25 (-15)	-20 (-5)
Viscosity index, min	. . .	75	. . .	. . .	. . .	. . .
Pour point, max, °C (°F)	-30 (-20)	-18 (-0.5)	-15 (5)	-35 (-30)	-30 (-20)	-23 (-10)
Stable pour point, max, °C (°F)	-30 (-20)	. . .	. . .	-35 (-30)	-30 (-20)	-23 (-10)
Flash point, °C (°F)	205 (400)	220 (430)	225 (435)	200 (390)	205 (400)	215 (420)

Source: Ref 11

(a) ASTM D 2602 modified.

(b) At -30 °C (-20 °F).

(c) At -20 °C (-5 °F).

(d) At -20 °C (-5 °F).

(e) At -15 °C (5 °F).

(f) At -10 °C (15 °F).

(g) At -5 °C (25 °F).

(h) ASTM D 3829 or CEC L-32-T-82.

(i) ASTM D 445.

(j) At -25 °C (-15 °F).

Service Classifications. Engine oil performance requirements are established by the API, ASTM, and SAE. These requirements are based on standardized tests and are described by the the API Engine Service Classification System, which was first introduced in 1969 and 1970. The API service symbol (Fig. 4), established in 1983, is devised to communicate engine oil quality and performance to the general public in order to help them select oils that meet manufacturer recommendations for use in the intended application (Ref 12). The upper part of the symbol displays the API service category; the center part, the SAE viscosity grade; and the bottom part, the energy-conserving feature, if applicable. Only licensees are authorized to display this symbol. The total number of such licensees at year-end 1990 in the U.S. and Canada was 210. In countries outside North America, there were 53 licensees.

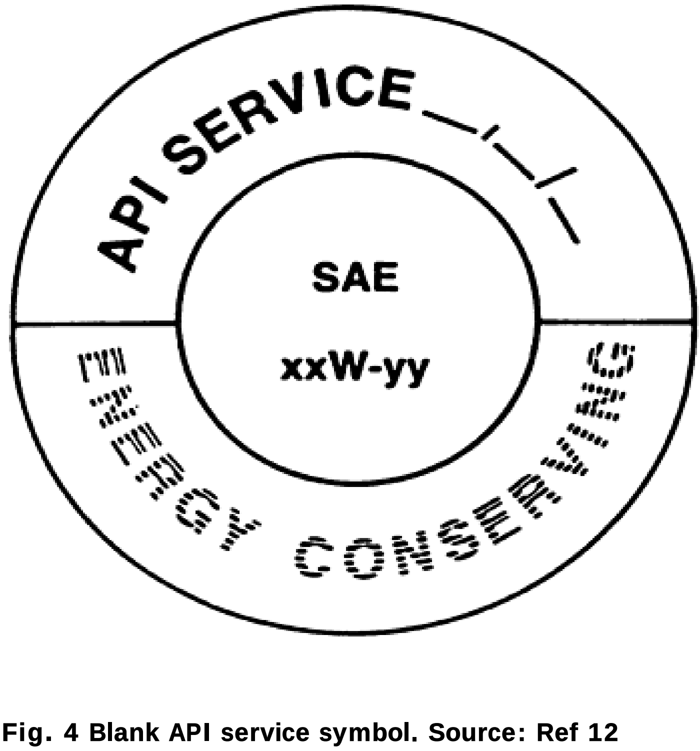


Fig. 4 Blank API service symbol. Source: Ref 12

API service categories for gasoline engine oils range from SA to SG, where S stands for service, and are used by service stations, garages, and the like. For diesel engine oils, the API categories range from CA to CE, where C stands for commercial. The most recently introduced API categories, SG and CF-4, are designed for the most severe operations (Ref 13). In general, oils that meet higher service requirements are suitable for use in the lower service class in the same category. The API SA to SF categories for gasoline engines and CA and CB categories for diesel engine oils are obsolete and are satisfied by API SG and CC (and higher service classes), respectively. Only the most recent categories appear within the symbol described above.

The SAE defines the need for a category, the ASTM develops or selects the test techniques, and the API develops the user language to define it. The engine oil classifications for gasoline engines are provided in Table 2, and those for diesel engines, in Table 3 (Ref 14). The performance requirements for some of these categories are available in Ref 15. The performance criteria measure protection against rust, varnish, deposits,sludge, wear, high-temperature oil thickening, and ring sticking. Figure 5 and 6 show the engine parts that are commonly rated.

Table 2 Engine oil classification systems for automotive gasoline engine services

API service categories	Previous API service categories	Related industry definition	Engine test requirements
SA	ML	Straight mineral oil	None
SB	MM	Inhibited oil only	CRC L-4[(a)] or L-38: sequence IV[(a)]
SC	MS (1964)	1964 MS warranty approved	CRC L-38; sequence IIA[(a)]; sequence IIIA[(a)] sequence IV[(a)], sequence V[(a)]; Caterpillar L-1[(a)] (1.0% sulfur fuel)
SD	MS (1968)	1968 MS warranty approved	CRC L-38; sequence IIB[(a)]; sequence IIIB[(a)], sequence IV[(a)]; sequence VB[(a)]; Falcon rust[(a)]; Caterpillar L-1[(a)] or 1H[(a)]
SE	None	1972 warranty approved	CRC L-38; sequence IIB[(a)] or IIC or IID; sequence IIIC[(a)], or IIID; sequence VC[(a)] or V-D
SF	None	1980 warranty approved	CRC L-38; sequence IID; sequence IIID; sequence V-D
SG	None	1989 warranty approved	CRC L-38; sequence IIID; sequence IIIE; sequence VE; Caterpillar 1H2
ILSAC	None	Proposed	CRC L-38; sequence IID; sequence IIIE; sequence VE; Caterpillar 1H2, sequence VI

Source: Ref 14

(a) This test is obsolete; engine parts, and/or test fuel, and/or reference oils are no longer generally available and the test is no longer monitored by the test developer or ASTM.

Table 3 Engine oil classification systems for automotive diesel engine services

API service categories	Previous API service categories	Related military and industry designations	Engine test requirements
CA	DG	MIL-L-2104A	CRC L-38; Caterpillar L-1 (0.4% sulfur)
CB	DM	Supplement 1	CRC L-38; Caterpillar L-1 (1.0% sulfur)
CC	DM	MIL-L-2014B: MIL-L-46152B	CRC L-38; sequence IID; Caterpillar 1H2
CD	DS	MIL-L-45199B, Series 3; MIL-L-2104C; MIL-L-2104D	CRC L-38; Caterpillar 1G2
CD-11	None	MIL-L-2104D	CRC L-38; Caterpillar 1G2; DD 6V53T
CE	None	None	CRC L-38; Caterpillar 1G2; Cummins NTC-400; Mack T-6; Mack T-7
CF[(a)]-4	None	None	CRC L-38; Caterpillar 1K; Cummins NTC-400; Mack T-6; Mack T-7

(a) New categories CF and CF-2 are being established.

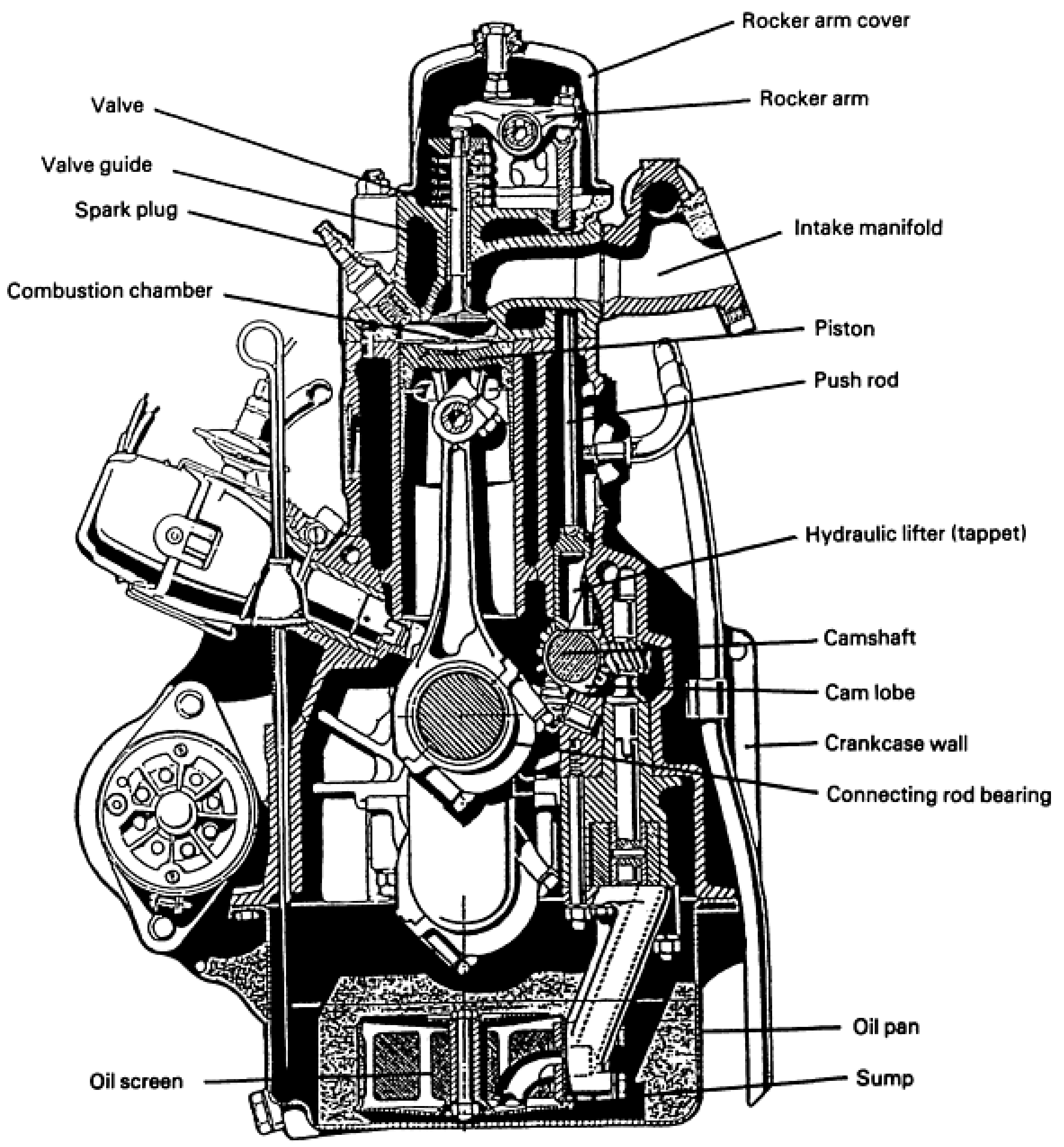


Fig. 5 Six-cylinder in-line engine

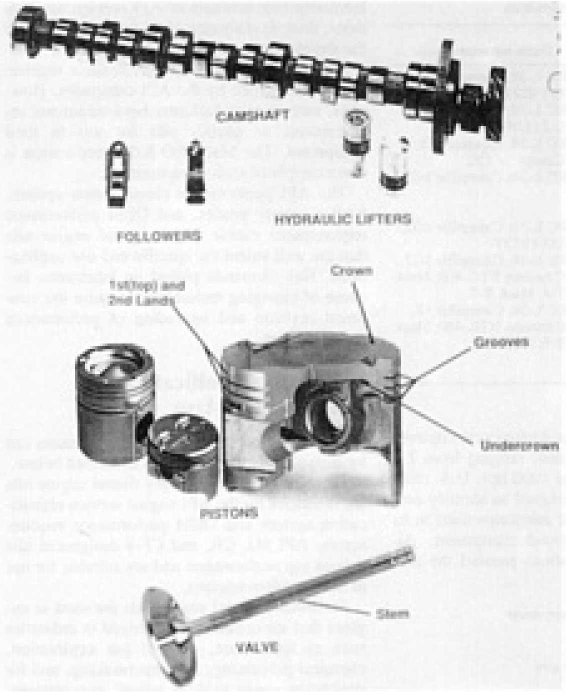


Fig. 6 Camshaft, followers, pistons, and valves

Military and OEM Requirements. The needs of U.S. military equipment differ from those of commercial equipment. The military operates large and diversified fleets of vehicles that contain two- and four-stroke cycle engines, air-cooled and liquid-cooled, ranging from 1.5 to over 745 kW (2 to over 1000 hp). U.S. military specifications are designed to identify performance requirements of lubricants used in its tactical and combat ground equipment. Although military specifications parallel the performance requirements of API service designations, their development does not always follow the development of the API categories.

Most OEMs accept the performance requirements established by the API categories. However, certain manufacturers have additional requirements to qualify oils for use in their equipment. The Mack EO-K/2 specification is one example of such a requirement.

The API performance classification system, SAE viscosity grades, and OEM performance requirements enable the design of engine oils that are well suited for specific end-use applications. New demands placed on lubricants, because of changing technology, require the continual revision and upgrading of performance specifications.

Lubricant Classification Based on End-Use

Depending on the application, lubricants can be classified by the categories described below.

Gasoline and heavy-duty diesel engine oils are identified by the API engine service classification system and OEM performance requirements. API SG, CE, and CF-4 designated oils deliver top performance and are suitable for use in most modern engines.

Stationary diesel engine oils are used in engines that are commonly employed in industries such as agriculture, oil and gas exploration, chemical processing, and papermaking, and for generating power to drive pumps, compressors, drilling equipment, and other auxiliary machinery. API CD and CD-II oils are often considered suitable for use in these engines.

Railroad diesel engine oils are typically SAE 40 viscosity grade for monograde oils and 20W-40 for multigrade oils, and are derived from paraffinic and naphthenic base oil bends. These oils have a sufficient base reserve (total base number of 10-20) to neutralize acidic combustion by-products. The total base number (TBN) is the acid-neutralizing power expressed in terms of mg of potassium hydroxide (KOH) per gram of the lubricant. These oils have a limit on the amount of zinc present (10 ppm maximum) because of the tendency of zinc-containing antiwear agents to attack silver

piston pin bushings formerly used in railroad engines. Most of these oils are formulated to meet API CD level performance, but the final approval depends on lubricant performance in a number of OEM-specified tests.

Marine diesel engine oils have two main categories: oils for low-speed engines and oils for medium-speed engines. Low-speed, crosshead engines are used for large ocean-going ships. Medium-speed, trunk-piston engines are usually used for ships of lesser size.

Low-speed two-stroke cycle engines use heavy fuel and are typically of the crosshead type. These engines are equipped with diaphragms and stuffing boxes, which separate power cylinders from the crankcase. This allows the use of different lubricants for the power cylinders and the crankcase (Fig. 7) (Ref 16).

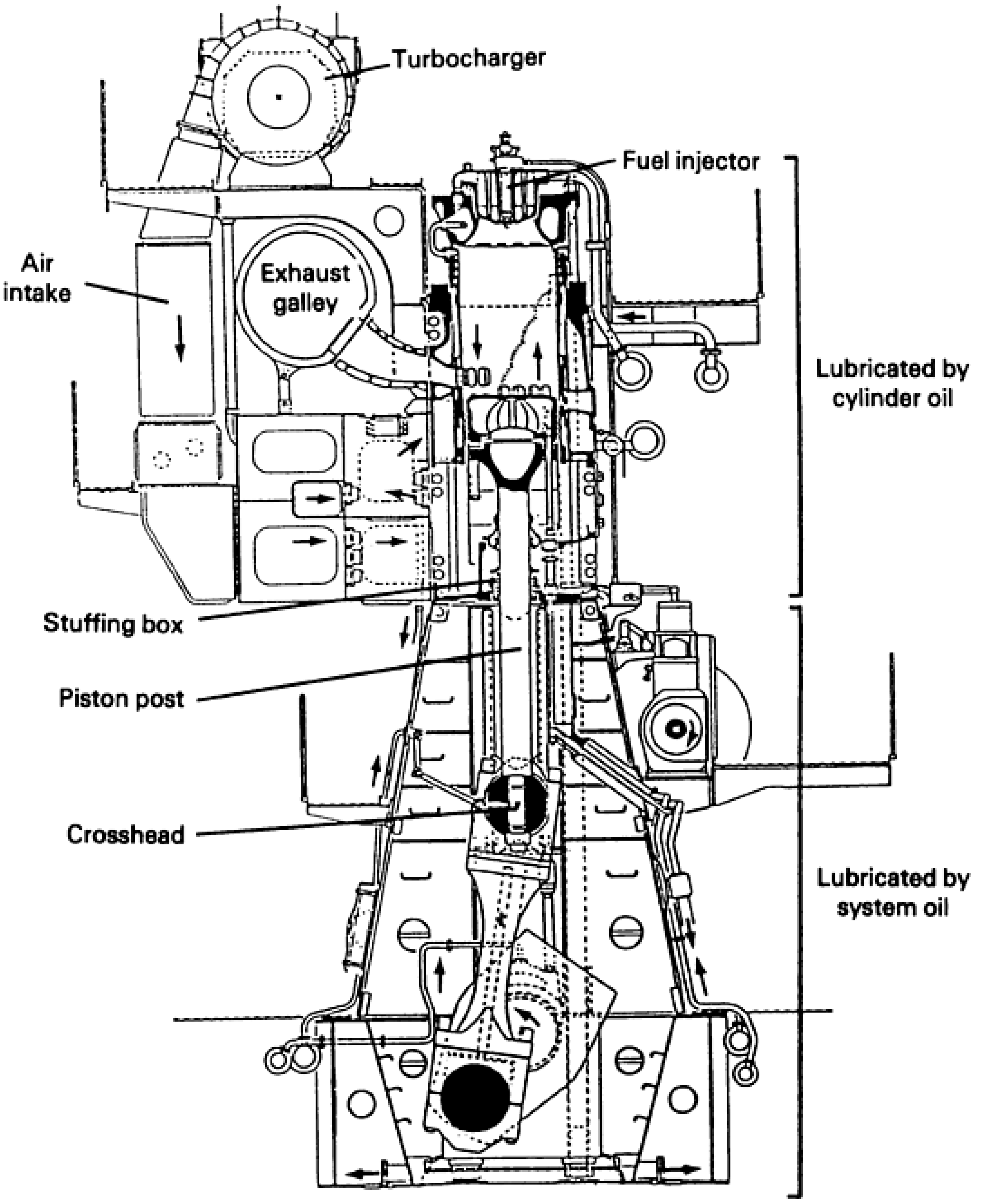


Fig. 7 Cross section of crosshead marine diesel engine. Source: Ref 16

The cylinder oil lubricates piston skirts and cylinders, and handles corrosive combustion products from heavy fuel. Therefore, it provides wear protection and engine cleanliness. Marine cylinder lubricants are typically SAE 50 viscosity grade, and have a suitable base reserve (TBN between 60 and 100) to neutralize the acidic combustion products arising from sulfur in the fuel. The cylinder oils are once-through lubricants because they are injected into the cylinder and burned with the fuel.

The crankcase lubricant lubricates bearings, gears, and other engine components. Crankcase or system oils are SAE 30 viscosity grade and are formulated to handle rust, oxidation, deposits, and wear. They typically have a TBN that ranges from 2 to 9.

Medium-speed trunk-piston engines have a single lubricating system and are run on distillate or heavy fuels. They use a lubricating oil of SAE 30 or 40 viscosity grade and have a TBN that ranges from 12 to 40, depending on the fuel quality. Again, the base reserve is necessary to neutralize acidic combustion products that are due to fuel sulfur. For marine lubricants, water tolerance is an additional desirable property that ensures that their lubricating ability will not suffer in case of water contamination. Lubricants with API CD performance are commonly used in this application.

Natural gas engine oils are used in stationary engines that operate on natural gas. Natural gas engines can be two-stroke or four-stroke cycle, and either naturally aspirated or turbocharged. In addition to stationary applications in the natural gas industry, these engines are used in petroleum refining, for power generation, and in agricultural irrigation service.

Lubricant performance is established by OEM specifications. The API CD oils are suitable for use in this application. The OEMs use sulfated ash to recommend oils for their equipment, with a preference for ash levels of 0.5% or less. Generally, the four-stroke engine manufacturers encourage the use of higher-ash oils than the two-stroke engine manufacturers. The SAE has determined the need for new engine categories for engine oils that are intended for use in this application and has requested that the API and ASTM define such categories and write user language.

Aviation engine oils are of two types: those used in piston engines and those used in turbine engines. The aircraft piston engines utilize engine oils to lubricate the crankshaft, connecting rods, pistons, rings, bearings, and other moving parts. Ashless dispersant oils that range from SAE 30 to 60 viscosity are used most widely.

Jet turbine engine oils are formulated using synthetic base stocks, usually polyol ester and ester types, which are used to lubricate metal surfaces whose temperatures are up to 675 °C (1250 °F). Their main function is to control deposit formation on the hot surfaces. A mineral oil based lubricant will not provide satisfactory performance at these high temperatures, because of insufficient stability. U.S. military specifications MIL-L-23699C and MIL-L-7808G are used to select lubricants for use in this application.

Two-stroke cycle engine oils generally pertain to lubricants designed for the SI (gasoline-fueled) two-cycle engines. Two-stroke cycle engine oils for the CI (diesel-fueled) engines were described in the section on marine diesel engine oils. Two-stroke SI engines are commonly used in marine outboard engines, motorcycles, mopeds, motor scooters, all-terrain vehicles, snowmobiles, and chain saws. The performance of two-stroke cycle engine oils is described by SAE service classifications (Ref 17) and National Marine Manufacturers Association TC-W II performance requirements.

The SAE miscibility and fluidity grades, described in SAE J 1536 (Ref 10), assess the ability of a lubricant to mix with the fuel at prescribed temperatures. Lubricant performance is assessed on the basis of its ability to protect against deposit formation, ring sticking, piston and cylinder scuffing, preignition, spark plug fouling, and exhaust port blocking. TC-W II specified oils are the most recommended oils for use in modern outboard cooled engines. A TC-W III specification is under consideration at this time.

Two-stroke cycle engine oils can be either ashless or ash-producing. Ashless oils dominate the U.S. two-stroke cycle engine oil market because of their preferred use in outboard engines. Ash-producing oils are more cost effective, and are widely used in nonoutboard applications.

Lubricant-Related Causes of Engine Malfunction

Most of the problems associated with internal combustion engine lubrication are related to the by-products of combustion, their entry into the crankcase through blowby, and lubricant decomposition. This implies that fuel and lubricant quality and combustion are both important parameters. The major causes of engine malfunction that are due to lubricant quality are deposit formation, lubricant contamination, oil thickening, oil consumption, ring sticking, corrosion, and wear.

Deposit Formation. The two main sources of lubricant contamination are blowby from the combustion chamber and gases and volatiles from the crankcase, which are vented into the intake manifold as an antipollution measure (Ref 18). The blowby is a mixture of nitrogen oxides (NOx), sulfur compounds (SO_2, SO_3, H_2SO_4), carbonyl compounds, hydrocarbons (unburned fuel), peroxides and radicals, air, water, and carbon dioxide. These species are blown past the piston rings and down the valve guides into the crankcase. There, they can interact with each other and the lubricant to form harmful products. Such products include soot, carbon, lacquer, varnish, and sludge (Ref 19).

Soot is an important particulate contaminant in crankcase lubricants. It can be derived from the inefficient combustion of fuel and the burning of lubricating oil that passes along pistons into the combustion chamber. Fuel-derived soot, which is more commonly encountered in diesel operation, is due to the use of fuel with a wide boiling range, which, under certain operating conditions, does not burn completely.

Soot is not pure carbon and contains an appreciable amount of hydrogen, oxygen, and sulfur in a combined form. Soot particles are basically hydrocarbon fragments with hydrogen atoms partly stripped off. These particles are strongly attracted to each other and to polar compounds in the oil, and tend to form aggregates. Soot deposits are soft and flaky in texture and are commonly found in the combustion chamber.

Carbon deposits, more prevalent in diesel engine operation, are hard and result from the carbonization of the liquid lubricating oil and the fuel on hot surfaces. Carbon deposits have lower carbon content than soot and, in most cases, contain oily material and ash. Carbon deposits are commonly found on the piston top lands and crowns, in the piston ring grooves, and on the valve stems.

Lacquer and Varnish. The unburned air-fuel mixture and either oxidized or partially oxidized reactive intermediates in the blowby promote lubricant oxidation. This results in the formation of a variety of oxygenated products which, when exposed to high temperatures, result in lacquer and varnish (Ref 20). The term lacquer usually describes this type of deposit in diesel engines, whereas varnish is used to describe such deposits in gasoline engines. Lacquer is often derived from the lubricant and is generally water-soluble. Varnish, on the other hand, is fuel-related and is acetone-soluble.

Lacquer is commonly found on pistons and cylinder walls, and in the combustion chamber. Varnish occurs on valve lifters, piston rings, and positive crankcase ventilation (PCV) valves.

Sludge. Three major causes of sludge formation are oxidation of the lubricant, oxidation and combustion products in the blowby, and the accumulation of combustion water and dirt. Sludge can vary in consistency from that of mayonnaise to that of a baked deposit. Heat can drive off water and thereby cause a change in consistency. Low-temperature sludge, more prevalent in gasoline engines, is more watery in appearance and forms at temperatures below 95 °C (205 °F). High-temperature sludge, more common in diesel engines, forms above 120 °C (250 °F). Sludge is commonly found in areas of low oil velocity, such as crankcase bottoms and rocker boxes.

Mechanism of Deposit Formation. The volume of deposits in an engine depends on the fuel quality, engine operating conditions (speed, load, and temperature), quality of combustion and blowby (its oxygen content, presence of sulfur and nitrogen compounds, and so on), and the integrity of the seal between the combustion chamber and crankcase.

Low-temperature deposits (soot, varnish, and low-temperature sludge) are usually encountered in gasoline engines subjected to intermittent (stop and go) driving, because this type of operation does not allow the engine to achieve the optimum temperature necessary to drive off contaminants (Ref 18).

High-temperature deposits (carbon, lacquer, and high-temperature sludge) are typically found in gasoline engines that experience long continuous operation and in diesel engines (Ref 20). These deposits result from thermal and oxidative degradation of the lubricant and additives. Additional stresses on the lubricant stem from the increased emphasis on cleaner exhaust emissions and efficient combustion of fuel at high temperatures; the use of power accessories, such as air conditioning; and body-style changes that affect cooling.

Nitrogen oxide in the blowby (a consequence of combustion) and high temperatures cause oxidative and thermal degradation of the lubricant. The result is the formation of oxygenated products that act as precursors to deposit-forming species. These products have 15 to 50 carbon atoms and contain hydroxy and carboxy functional groups. They are polyfunctional molecules with the ability to polymerize thermally to form higher molecular weight products.

If the oxygen content of the precursors is low, and the polymer product is of low molecular weight and good oil solubility, then only oil thickening is observed. However, if the oxygen content of the precursors is high, and the polymerization results in products of low lubricant solubility, then varnish and resin are formed. When the polymerization occurs on hot metal surfaces, varnish is formed, and when it occurs in the bulk lubricant, resin is formed.

Because of its low solubility in the lubricant, resin tends to separate out on metal surfaces as lacquer, which can be amber or black in color. At low soot levels, the interaction of resin and soot results in resin-coated soot particles, which separate out on piston surfaces as black lacquer. At high soot levels, extensive interaction between soot and resin forms soot-coated resin particles that have little or no ability to adhere to metal surfaces. Instead, they accumulate as deposits in areas of low oil flow, such as in the grooves behind the piston rings (Ref 20). Sludge results from the interaction of the oxygenates with soot in the presence of oil and water.

Deposit levels depend on fuel quality. In the case of sulfur-rich fuels, the oxidation of sulfur results in the formation of sulfur acids, which catalyze the rearrangement of hydroperoxides to carbonyl compounds and their subsequent polymerization to resin. The addition of basic detergents, which are used in diesel engine oils (for marine and railroad use) to counter the adverse effects of sulfur, helps to alleviate this problem.

Oil thickening can result from a combination of oxidative degradation of the lubricant and the accumulation of insolubles (Ref 21). The auto-oxidation of the lubricant, accelerated near the oxidation inhibitor depletion stage, can lead to oxygenated products, which can cause a viscosity increase through polymerization (Ref 20). The contaminant-related thickening arises from the suspension of fuel-derived insolubles in the bulk lubricant.

Oil consumption is related mainly to the lubricant that travels past the piston rings and valves and burns in the combustion chamber. Burning of the lubricant, along with inefficient fuel combustion, leads to soot and carbon deposits on the inside of the combustion chamber, piston top land, ring grooves, and so on (Ref 19). The extent of oil consumption depends on a number of lubricant and equipment design-related factors. The viscosity, volatility, and seal-swell characteristics of the lubricant play an important role in this regard.

A certain minimum amount of oil is necessary for proper lubrication of the cylinder walls and pistons. High oil consumption is indicative of problems in pistons and cylinders, such as increased wear of the cylinder, bore polishing, stuck piston rings, or out of square grooves (Ref 4). Under these circumstances, the blowby gases have an increased chance of entering the crankcase, which is likely to further complicate the situation.

There is some evidence that oil consumption contributes toward particulates in diesel emissions. A consolidated effort is being made by OEMs and the lubricant marketers in the United States, Europe, and Japan to minimize these particulates, because of environmental concerns. One of the strategies to achieve this goal is to use lower-thickness oil films. This strategy is quite effective, except that beyond a certain limit, the oil film thickness will be ineffective for proper lubrication, and ring/liner scuffing problems will occur. Lubricant volatility is another important factor responsible for increased oil consumption. Lighter base oils that are used to formulate multigrade diesel oils not only contribute to the formation of a less-effective lubricating film, but also can readily leak past the piston rings and burn (Ref 22).

Ring Sticking. The major cause of ring sticking is the formation of deposits in the piston grooves. The consequence is a loss of oil seal, which not only favors blowby, but also results in poor heat transfer from the piston to the externally cooled cylinder wall. This is quite serious, because thermal expansion of the pistons will lead to loss of compression and, ultimately, to engine seizure.

Corrosion and Wear. Diesel fuel with a high sulfur content causes piston ring and cylinder wear, especially in large, slow-speed, marine diesel engines. Corrosive wear, more commonly associated with combustion and oxidation products, results from the attack of either sulfur acids or organic acids on ferrous surfaces. This happens when the engine operating temperatures are below the dew point of these acids. The organic acids can originate from thermo-oxidative decomposition of the lubricant or the additive system. This kind of wear is controlled by using lubricants that have a base reserve.

Lubricant Formulation

The inherent ability of mineral oil to act as a lubricant is not sufficient for most modern applications. Synthesized chemicals, called additives, are blended in base stocks, such as mineral oils, to help them perform effectively in demanding lubricating environments. A finished lubricant comprises a base fluid and a performance package, and, in the

case of multigrade oils, a viscosity modifier. The performance package can constitute as much as 30% of the total composition, depending on the desired performance level and the severity of end-use requirements.

Base fluids, or lubricant base stocks, are either of mineral or synthetic chemical origin. Mineral oil base stocks are obtained directly from petroleum fractionation, whereas synthetic base stocks are manufactured through transformations of petroleum-based organic chemicals. Partially synthetic base stocks are compatible mixtures of mineral oil and synthetic base stocks.

Mineral oil base stocks (lubricating oils) are obtained from the crude oil fraction, which boils at temperatures greater than 340 °C (645 °F) through a series of separation processes. These processes include vacuum distillation, solvent extraction, and wax removal. Some base stocks are hydrotreated or hydrocracked to improve color, oxidation resistance, thermal stability, and viscosity (Ref 23).

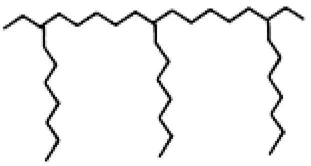
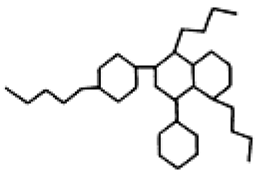
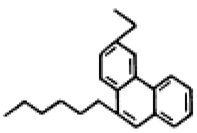
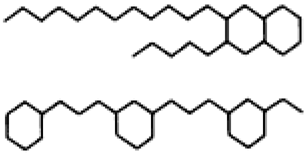
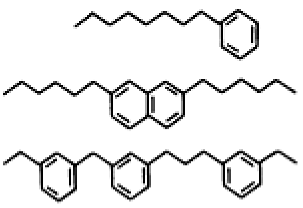
The more highly refined oils have little or no sulfur, and are therefore more prone to oxidation. The oxidation resistance of oils, which are more susceptible to oxidation because of the presence of unsaturation, can be greatly improved through hydrogenation. Most automotive and industrial lubricants, as well as greases, use solvent-extracted base oils. Both solvent-extracted and hydrotreated base oils are primarily used for premium products, such as turbine, hydraulic, and circulating oils.

Mineral oil bases stocks typically contain hydrocarbon molecules that contain from 20 to 70 or more carbon atoms. These base stocks are sometimes characterized as being a paraffinic, naphthenic, or aromatic type, depending on the nature of the hydrocarbons present. Paraffinic base stocks contain long saturated hydrocarbon chains; naphthenic base stocks contain five- or six-membered saturated cyclic rings; and the aromatic base stocks contain unsaturated cyclic (aromatic) rings. In reality, none of the base stocks are composed of a single structural type. Instead, they contain mixed structures, with one or the other predominating. Paraffinic base stocks, for example, contain all three types of structures, but paraffinic structures are present in the largest amount. The relative distribution of each type of hydrocarbon structure has a significant effect on the viscosity-temperature relationship and the pour point of the base stock.

The viscosity-temperature relationship of an oil is expressed in terms of its viscosity index (VI), which indicates the degree of change in viscosity of an oil with a change in temperature within a given temperature range. A high VI signifies a relatively small change, and a low VI, a large change. The VI of an oil can be determined by measuring the kinematic viscosity of the oil at 40 °C (105 °F) and 100 °C (212 °F), and by using ASTM tables (Ref 24).

Paraffinic base stocks are generally preferred because of their high viscosity index (HVI). However, because they contain a substantial amount of wax, they have high pour points. Naphthenic base stocks, which have lower viscosity indices, are preferred only in certain applications, because of their lower wax content and lower pour points. Aromatic stocks are the least preferred, because of their poor oxidation resistance, tendency to form black sludge at high engine operating temperatures, low VI character, and their suspected carcinogenic nature. The relationship between various lubricant properties and structures is shown in Table 4.

Table 4 Relationship between properties and hydrocarbon structures

Hydrocarbon types		Main properties
Straight chain paraffin		Viscosity varies little with temperature; good oxidation resistance; high pour point
Branched chain paraffin		Viscosity varies little with temperature (except when divided); good oxidation resistance; may have low pour point
Naphtha rings with short paraffin side chain		Good oxidation resistance; low pour point; viscosity varies greatly with temperature; becomes pseudoplastic under cool conditions
Aromatic rings with short paraffin side chain		Pour point varies according to structure; good thermal stability; viscosity varies greatly with temperature; easily oxidizes
Naphtha rings with long paraffin side chain		Viscosity varies little with temperature; good oxidation resistance; may have low pour point
Aromatic rings with long paraffin side chain		Viscosity varies little with temperature; may have good oxidation resistance if cycles are not numerous; may have low pour point

Source: Ref 25

Source: Ref 25

Synthetic base stocks are manufactured by chemical reactions such as alkylation, polymerization, and esterification. Commercial synthetic fluids include synthetic hydrocarbons (alkylaromatics and polyolefins), organic esters (dibasic acid esters, polyol esters, and polyesters), and miscellaneous other organic materials (phosphate esters, polyalkylene glycols, and polyphenyl ethers) (Ref 26, 27, 28, 29). Polyolefins, dibasic acid esters, polyol esters, and polyalkyleneglycols (polyether glycols) are the most widely used synthetic lubricants. Their chemical structures and principal applications are shown in Fig. 8 (Ref 25).

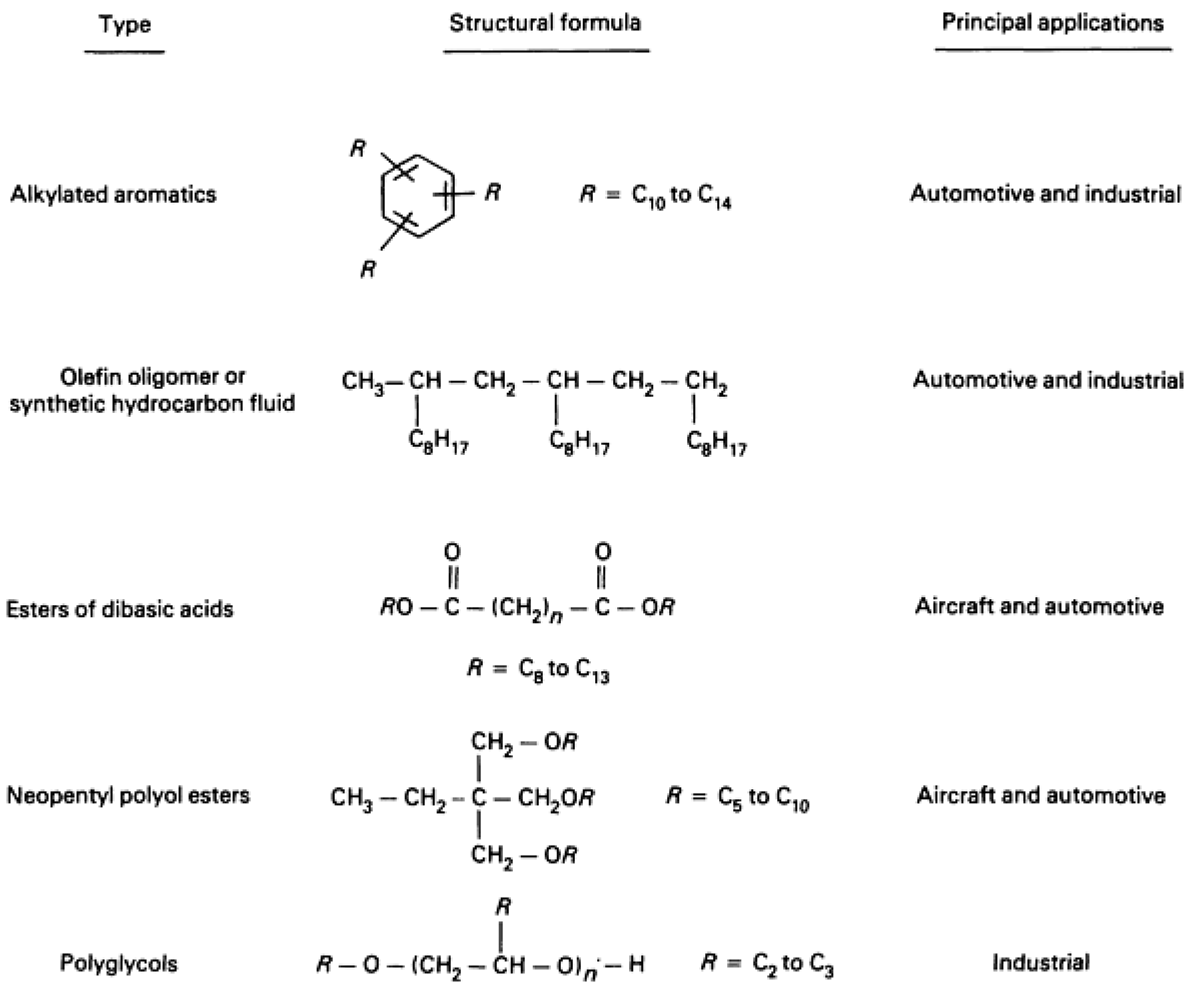


Fig. 8 Synthetic base stocks and their uses. Source: Ref 27

Synthetic base stocks are used in situations where mineral oil base stocks do not provide satisfactory performance, because of their inherent limitations. Synthetic base stocks can be devised to possess any number of unique and desirable properties. Commonly sought properties include enhanced thermal stability, high viscosity index, fire resistance, low pour point, low volatility, and low toxicity (Ref 28).

Table 5 compares some physical properties of hydrofinished HVI stocks and synthetic base stocks. The levels and types of additives used in either synthetic or partially synthetic base stocks (blends of synthetic base stocks and mineral oils) are similar to those used in mineral oils.

Table 5 Physical properties of hydrofinished high viscosity index (HVI) stocks and synthetic base stocks

Type	Kinematic viscosity, mm^2/s			Viscosity index	Pour point		COC flash	
	At 40 °C (105 °F)	**At 100 °C (212 °F)**	**At -40 °C (-40 °F)**		**°C**	**°F**	**°C**	**°F**
Hydrofinished HVI stocks								
90 neutral	17.40	3.68	. . .	92	-15	5	190	375
100 neutral	20.39	4.11	. . .	101	-13	9	192	378
200 neutral	40.74	6.23	. . .	99	-20	-5	226	440
350 neutral	65.59	8.39	. . .	97	-18	-0.5	252	485
650 neutral	117.90	12.43	. . .	96	-18	-0.5	272	520
150 bright stock	438.00	29.46	. . .	95	-18	-0.5	302	575
Synthetic base stocks								
Alkylated aromatic	26.84	4.99	9047	119	-18.3	-0.9	224	435

Olefin oligomer	16.77	3.87	2371	126	-43.0	-45	221	430
Ester of dibasic acid dioctyl sebacate	18.2	. . .	3450	176	-15.6	4	232	450
Ester of trimethylol propane (C_7)	13.94	3.4	2360	. . .	<-18.3	<-0.9	232	450
Polyglycol	45.69	. . .	. . .	150	-21	-6	177	350

Source: Ref 27

Performance Package

The performance package contains a number of chemical compounds (additives) that are blended in base stocks to make them suitable for use as lubricants.

The quality and quantity of additives in the performance package depend on the properties of the base fluid and the intended application. The functions of a lubricant are to limit and control friction, metal to metal contact, overheating (from friction and combustion), wear, corrosion, and deposits. The specific properties that the lubricant should have in order to perform these functions include suitable viscosity, slipperiness, high film strength, low corrosivity, low pour point, good cleansing and dispersing ability, nontoxicity, and low flammability. In addition, the lubricant should not foam, that is, it should be capable of getting rid of air (oxygen) to minimize oxidation and maintain its lubricating characteristics.

The viscosity of the oil should be such that it can provide proper lubrication, both at low temperatures and at the high temperatures encountered during engine operation. This implies that oil viscosity at low temperatures should not interfere with starting and lubrication during the warm-up period, and that loss of viscosity at high temperatures is minimal. As a general rule, large clearances and high loads demand high viscosity, and high speeds demand low viscosity.

The oiliness and film strength of the lubricant are related to its ability to reduce friction and wear under boundary lubrication conditions and to protect against corrosion. The oil itself should not be corrosive, and it should provide protection against corrosion.

The pour point is one of the means used to determine the flow characteristics of lubricants at low temperatures. Lubricants that contain wax will provide little or no lubrication below their pour point, because wax crystals can separate and cause blockage of the oil pump inlet.

Detergency is the property related to the ability of a lubricant to keep engine internal surfaces clean through film formation and by neutralizing the acidic products of combustion and oxidation. If not controlled, these products will promote the formation of engine deposits. Dispersancy is the ability of the lubricant to suspend harmful products through association.

The lubricant should resist oxidation in order to minimize the formation of harmful products, such as acids, lacquer, and sludge. These products will adversely affect engine performance.

These functions are carried out by additives such as dispersants, detergents, antiwear agents, friction modifiers, oxidation inhibitors, corrosion inhibitors, pour-point depressants, foam inhibitors, and viscosity improvers. Descriptions of these additives and their functions are provided in Ref 30 and 31, as well as in the article "Lubricant Additives and Their Functions" in this section of this Volume.

Performance Testing

All engine oils are performance tested prior to being introduced to the marketplace. Trade associations and technical societies that contribute to testing and approval processes in the United States are the API, ASTM, Chemical Manufacturers Association, Motor Vehicles Manufacturers Association, SAE, and the National Marine Manufacturers Association. The Coordinating European Council, in western Europe, and the Japanese Automotive Standards Organization and Society of Automotive Engineers, Japan, establish the performance requirements in those countries. The need for a new performance level may relate to new engine design, current equipment problems, or future goals.

As a preliminary step, a number of oils with demonstrated field performance are selected as reference oils. These oils are run in the laboratory using production engines from various manufacturers simulating actual service conditions. The engine parts are periodically examined to closely monitor the performance of the lubricant. A good correlation between field performance and laboratory performance standardizes the test to evaluate new lubricants.

The development of laboratory engine tests is costly and time consuming. Therefore, future performance requirements have to be anticipated and defined as accurately as possible. This can only be done through the cooperation of equipment manufacturers, lubricant and additive marketers, and the end users. The laboratory engine tests are expensive to house, operate, and maintain. Despite continued efforts to devise tests to evaluate lubricant performance quickly and inexpensively (Ref 25), no good substitute for engine tests has yet been found.

A number of key processes make up a quality test. These include the quality of parts, engine build, operating procedures, and test ratings. Low part-to-part variability is extremely critical to proper evaluation of the lubricant. If parts vary in quality, then it is difficult to assess whether the test results reflect the effectiveness of the lubricant or part-to-part variance. The way an engine is assembled also has an impact on engine test results. Therefore, it is important that the test engines are built according to an established procedure and with parts of good and consistent quality.

The ASTM has the prime responsibility of ensuring that the defined operating procedures for a particular test are followed. Most modern testing facilities use computer data acquisition and control systems to monitor the operation of a test. Ratings are conducted by qualified raters, who are trained by a self-policing industry-supported group. Previously, much emphasis was placed on the operations and ratings of a test. That emphasis is being supplanted by the realization that all four elements are equally important when conducting a quality test.

Several controlling bodies ensure that the established test procedures are properly carried out, and that the ratings are consistent among industry laboratories. In the United States, such bodies include the Test Monitoring Center (TMC), the ASTM surveillance panels, and the SAE Lubricants Review Institute (LRI).

The TMC, a key part of the ASTM monitoring system, is a neutral body that deals with the calibration and monitoring of test facilities to assure that the engine tests are conducted according to the procedures described by ASTM (Ref 32). The TMC also supplies the reference oils, validates stands for reference purposes, and approves laboratories for testing. If a laboratory is not operating according to the established guidelines, in the opinion of the TMC, then the TMC has the authority to invalidate it. Because the TMC oversees most test laboratories, it provides statistics and identifies trends.

The surveillance panels are composed of experts, usually from the test laboratories, who monitor a specific test and make sure that the test runs as prescribed and is providing consistent and on-target results.

The LRI ensures that the candidate products meet U.S. military performance requirements. In order to obtain approval, the additive or lubricant supplier submits the selected parts from the prescribed tests, along with test parameters and test results. If these are acceptable, a recommendation for approval is granted for the pertinent military specification. Final approval is granted by the military upon review of the LRI recommendation and supporting documentation. Because many candidate products also meet the requirements described by SAE J 183 and are marketed in the private sector, the results of the LRI recommendations also impact commercial oils (Ref 33).

Once the lubricants have met all the requirements established by SAE, API, and ASTM, some lubricant and additive marketers choose to field test their products. This is the final test of lubricant performance. It involves testing in fleets and can take one year or more. The test employs commercial vehicles or vehicles that involve extensive use. The engines in these vehicles are examined periodically to assess lubricant performance (Ref 34). A lubricant with demonstrated good performance in laboratory engine tests and field tests is ready for the marketplace.

References

1. E.F. Obert, *Internal Combustion Engines and Air Pollution,* Intext Educational Publishing, 1968, p 1-25
2. F.X. Sieloff and J.L. Musser, "What Does the Engine Designer Need to Know About Engine Oils?," paper presented to Detroit Section of the Society of Automotive Engineers, Mar 1982
3. E.F. Obert, *Internal Combustion Engines and Air Pollution,* Intext Educational Publishing, 1968, p 669-673
4. W.B. Chamberlin and J.D. Saunders, Automobile Engines, in *CRC Handbook of Lubrication,* Vol 1, CRC Press, p 3-44
5. L.F. Schiemann and J.J. Schwind, Fundamentals of Automotive Gear Lubrication, *Fuels and Lubricants Technology: An Overview,* SP.603, Society of Automotive Engineers, Oct 1984, p 107-115
6. E.L. Pilon, Gear Lubrication--I, *Lubrication,* Vol 66, 1980, p 1-12
7. N.P. Suh and N. Saka, Ed., *Fundamentals of Tribology,* MIT Press, 1980

8. L.D. Wedeven, What is EHD?, *J. Am. Soc. Lubr. Eng.*, Vol 31, June 1975, p 291-296
9. "Engine Oil Viscosity Classification," SAE J 300, Society of Automotive Engineers, June 1989 (revised)
10. "Two-Stroke-Cycle Engine Oil Miscibility/Fluidity Classification," SAE J 1536, Society of Automotive Engineers, Oct 1988 (revised)
11. "Ready Reference for Lubricant and Fuel Performance," Publication 1288 240-94R, Lubrizol Corporation
12. "Engine Service Classification System and Guide to Crankcase Oil Selection," Publication 1509, American Petroleum Institute, Mar 1985
13. "Engine Oil Performance and Engine Service Classification (Other than 'Energy Conserving')," SAE J 183, Society of Automotive Engineers, June 1989 (revised)
14. R. Muller and S. Papageorge, Engine Oil Performance Tests in the United States, *Lubrication,* Vol 74, 1988, p 1-29
15. "U.S. Gasoline and Diesel Engine Oil Specification," technical publication, Lubrizol Corporation
16. G.J. Hellingman and S. Barrow, "Shipboard Investigations with Selected Fuels of Tomorrow," paper presented at 14th International Congress on Combustion Engines (Helsinki), 1981
17. "Two-Stroke-Cycle Gasoline Engine Lubricants: Performance and Service Classification," SAE J 2116, Society of Automotive Engineers, 1990
18. K.L. Kreuz, Gasoline Engine Chemistry as Applied to Lubricant Problems, *Lubrication,* Vol 55, 1969, p 53-64
19. C.A. Bouman, *Properties of Lubricating Oils and Engine Deposits,* McMillan and Co., London, 1950, p 69-92
20. K.L. Kreuz, Diesel Engine Chemistry as Applied to Lubricant Problems, *Lubrication,* Vol 56, 1970, p 77-88
21. M.J. Covitch, B.K. Humphrey, and D.E. Ripple, "Oil Thickening in the Mack T-7 Engine Test--Fuel Effects and the Influence of Lubricant Additives on Soot Aggregation," paper presented at SAE Fuels and Lubricants Meeting, Society of Automotive Engineers (Tulsa), Oct 1985
22. P.M. Lewinsky, V.B. Cooke, and C.A. Andrews, "Lubrication Oil Requirements for Low Emission Diesel Engines," paper presented at AVL Conference on Engine and Environment (Graz, Austria), Aug 1989
23. G.K. Schmidt and E.J. Forster, Modern Refining for Today's Fuels and Lubricants, *Fuels and Lubricants Technology: An Overview,* SP.603, Society of Automotive Engineers, Oct 1984, p 1-10
24. "Viscosity Index Tables for Celsius Temperatures," ASTM Data Series Publication DS-39B, ASTM, Sept 1965
25. A. Schilling, *Motor Oils and Engine Lubrication,* Scientific Publications, Great Britain, 1968
26. "Physical and Chemical Properties of Engine Oils," SAE J 357, SAE Information Report, Society of Automotive Engineers, May 1984 (revised)
27. P.A. Asseff, "Lubrication Theory and Practice," Publication 183-320-59, Lubrizol Corporation
28. D.E. Ripple and J.F. Fuhrmann, "Performance Comparisons of Synthetic and Mineral Oil Crankcase Lubricant Base Stocks," paper presented at The Fourth International Colloquium (Esslingen, Germany), Jan 1984
29. E.R. Braithwaite, *Lubrication and Lubricants,* Elsevier Publishing, 1967, p 166-196
30. W.C. Gergel, "Lubricant Additive Chemistry," paper presented at The International Symposium on Technical Organic Additives and Environment (Interlaken, Switzerland), May 1984
31. J.F. Ford, Lubricating Oil Additives--A Chemist's Eye View, *J. Inst. Petrol.,* Vol 54, July 1968, p 188-210
32. C.C. Colyer, The ASTM Test Monitoring System--Purpose and Operation, *ASTM Standardization News,* Oct 1981, p 8-12
33. C.C. Colyer, "The SAE Lubricants Review Institute--Purpose and Operation," paper presented at the International Congress and Exposition (Detroit), Feb 1983
34. C.C. Colyer, K.E. Davis, M.E. Desing, S.E. Jaynes, and W.G. Copan, "Field Testing: The Final Proof of Performance," paper presented at the XXII FISITA Congress (Dearborn, MI), Sept 1988

Introduction to Wear

K.C. Ludema, University of Michigan

PERHAPS THE BIGGEST CHALLENGE in solving wear problems is that of anticipating the type(s) of wear to which components will be subjected. Material can be removed from a solid surface in only three ways: by melting, by chemical dissolution, or by the physical separation of atoms from the surface. The last method can be accomplished either by the one-time application of a high strain or by cyclic straining at lower magnitudes. Mechanical and chemical processes may operate separately or together, such as abrasion in a corrosive medium.

There is no organized catalog containing exact descriptions of the states of stress or chemical conditions imposed on materials subjected to wear over a large range of applications because the mechanics of sliding and other forms of wear have not been rigorously defined for all cases. However, a great deal of information is available on the wear response of materials within smaller ranges of conditions.

Lacking a first-principles means to select and match materials to any wear situation, a combination of failure analysis and simulative testing is required. The articles in the Section "Wear Monitoring and Diagnosis" of this Handbook describe the tools of wear analysis and provide guidance on how to observe the wear processes in the proper scale and with the proper perspective. The articles in the other Sections on Wear address various wear types, describing their mechanisms and mitigation. The classification of wear types is somewhat arbitrary, but all the major environments in which wear commonly occurs are covered.

The terms used in the following articles to describe wear vary based on the author's background and on the industry in which he works. Many terms are subjective to some extent. For example, we often refer to a sliding environment as "abrasive" when we feel gritty substances on or near the wearing surfaces. In fact, the wearing materials may not have been abrasively worn at all, but instead by some other process, such as by the chemical action of the environment. The reader is referred to the Glossary for more formal definitions of many of the terms associated with wear; however, not all terms and colloquial jargon are contained there. Instead, their context will be a guide to their meaning.

The major creative act in designing for wear resistance still lies with the designer. The articles in this section can serve only as general guidelines. They have been written mostly by research-oriented authors who study specific wear mechanisms by experiments that attempt to isolate only the mechanism of interest. Practical devices almost always wear by a sequence or combination of wear mechanisms. Neither designers nor researchers can anticipate which combinations will occur without experience from actual service. The wise course, then, is simulative testing, provided that one is armed with a solid knowledge of mechanisms and equipped with the proper perspective for observing what is going on.

Finally, because we cannot be certain that any given wear part will be used in a manner that we might anticipate, we apply methods of monitoring and diagnosis to enhance our powers of observation. Many of these issues are discussed in more detail in the articles that follow.

Surface Damage

Sture Hogmark, Staffan Jacobson, and Olof Vingsbo, Uppsala University (Sweden)

Introduction

THE TYPES OF DAMAGE that surfaces display as a result of tribocontacts will be surveyed in this article. Surface damage in this context is defined as topographical or microstructural changes, or both, in a surface layer. Micrographs of various sorts will obviously have a dominant role in this description, which can be thought of as an introduction to "tribography," defined as a micrographic description of topographical and structural modifications in tribocontacts. Diagnosing surfaces with respect to the mechanisms and processes that caused the damage, and relating those mechanisms to the geometry, materials properties, and other characteristics of the tribosystem, can be instrumental in reaching a deeper understanding of tribological components. The final objective of tribographical efforts is to achieve better tribosystems.

Surface damage to a tribosystem is most often generated in many consecutive small steps by a number of different micromechanisms that are active in the tribosurface. The observed damage is thus the cumulative effect on a macroscale of these micromechanisms. The present tribographical approach is based on a direct classification of surface damage features, as displayed on the resulting surface, rather than on the traditional classification by wear mechanisms that were active in producing it.

Classes of Surface Damage

In principle, a tribosurface may exhibit damage of a single type, but generally the pattern is a combination of two or more types. Examination and interpretation may be quite complex, but a classification into the idealized types illustrated in Fig. 1 is convenient and will be used in this article.

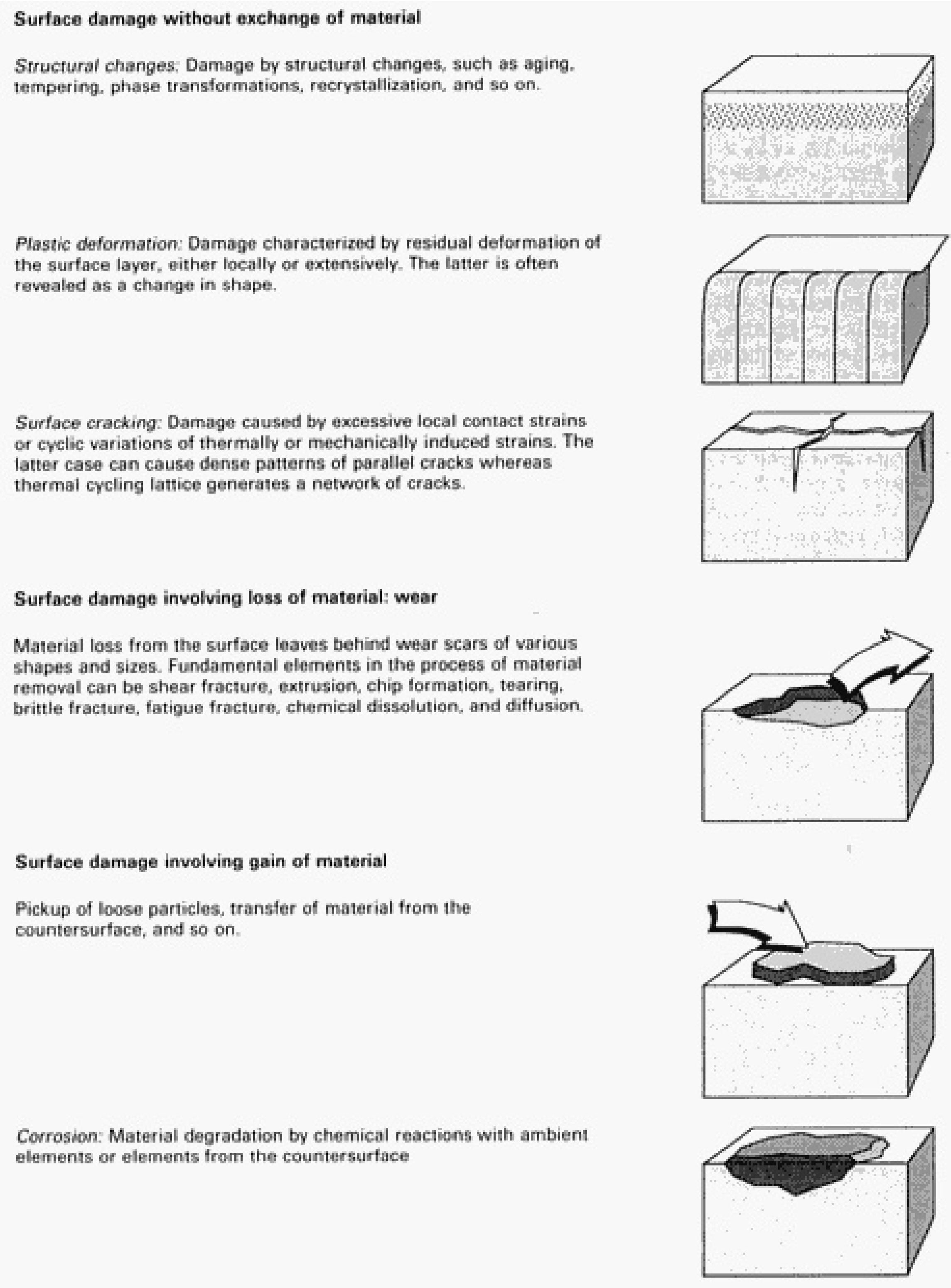


Fig. 1 Classification of surface damage

All the types of damage shown in Fig. 1 will in practice influence a surface layer. It can be noted that, using the present definition--topographical or structural changes of a surface layer--damage is not necessarily harmful. The loss of material and plastic deformation involved in the "running-in" of an engine naturally "damages" the virgin tribosurfaces, but is still favorable to the engine. Similar examples will be presented. It is obvious from Fig. 1 that loss of material is only one type of surface damage, but it is frequently the most important.

Interaction Between Different Types of Damage. The fact that the damage observed on a tribologically loaded surface normally is a combination of two or more of the types illustrated in Fig. 1 indicates that several mechanisms have been operating simultaneously. The overlapping mechanisms generally do not operate independently, but interact to mutually increase the effects of one another. An example of such synergistic interaction is the simultaneous corrosion and wear involved in high-temperature erosion. The oxidized layer is more prone to erosive wear, which in turn increases the oxidation rate by exposing bare metallic surfaces to the atmosphere, and so on.

Occasionally, however, the combined effects can lead to a net reduction of the individual effects. One example of such a counteractive interaction is the suppression of severe adhesive wear of metals in sliding contacts by the formation of an oxide layer. The oxide layer reduces the adhesive forces between the sliding bodies, which diminishes the probability of shear fracture of the metal surfaces. In favorable cases, the friction is reduced and the adhesive metallic wear is replaced by a mild wear of the slowly growing oxide layer.

Diagnosis of Surface Damage

Properties of the Surface Layer. Figure 1 demonstrates that, although surface studies render much information, they must be supplemented with microstructural studies on cross sections to give a more complete picture. The structure and even the composition of the damaged surface layer are often very different from those of the virgin material. Hence, it would not be relevant to consider the properties of the virgin material even if measured in the actual state of local stress and temperature.

The designation "active layer" refers to the surface layer with its composition, structure, thermal, and mechanical properties *during* the tribological contact, whereas the "relaxed (post-worn) layer" is the same layer with properties reached after unloading and cooling to ambient temperature (see Fig. 2). An important part of tribography is to study the relaxed layer, deduce its active state, and from there draw conclusions about mode of failure, improved design, materials selection, or other aspects depending on the purpose of the investigation.

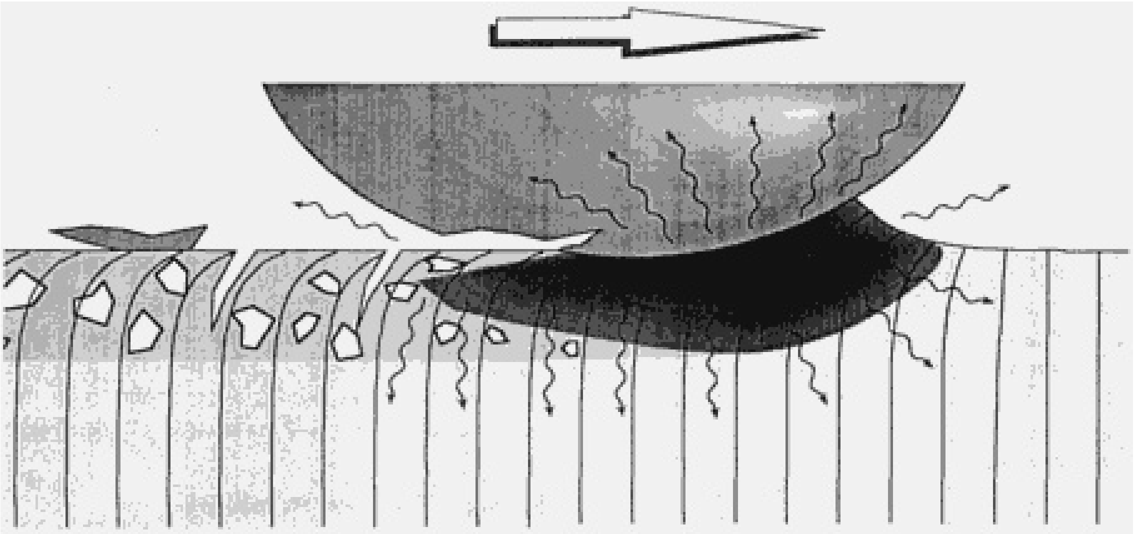


Fig. 2 Schematic of a single triboevent generating an active layer by transforming the original surface material under the influence of high normal and shear forces and frictional heat. Relaxation of this layer, involving additional phase transformations, will occur some time after the contacting asperity has passed.

Effects of the Surface Damage. When diagnosing a damaged surface by tribographical examination, it is important to be attentive to the type of problem the damage is causing--that is, to identify which type of damage sets the "life-limit criterion." The effects of a certain type of damage and the recommended action for reducing the problem vary greatly from case to case. For consumable wear parts, the only action to be taken should be to try to reduce the dominant wear mechanism, whereas there is no need to worry about transfer particles and so forth. When the requirements on the topography are high, however, as in rolling bearing surfaces and printing cylinders, small scratches, transfer particles, or even surface cracks may be detrimental to the function of the system. Generally, the problems can be divided into the following groups:

- Material losses eventually consume the wear part (for example, excavator teeth, shoe soles)
- Material losses deteriorate the function (by causing cutting edge blunting, excessive bearing clearance, and so on)
- Disadvantageous topography, structure, or composition deteriorates the function (for example, rough topography of forming tools or on optical lenses)
- Wear particles deteriorate the function (by causing abrasive wear, by contaminating end products such as in food processing, and so on)

There are also numerous examples of surface damage that may be highly beneficial to the function. During the running-in phase of many machines, for example, tribosurfaces are smoothed by wear and plastic deformation, which leads to improved conformity, reduced stress concentrations, and enhanced lubrication. A further mechanism is the formation of wear-protective overlayers, such as those frequently formed on cutting tool edges or on hot rolls. Friction-reducing layers also are sometimes formed on the tribosurface--in dry sliding between polymers and metals or in lubricated systems with oils containing chemically active additives, for example.

Relative Importance of Rival Wear Mechanisms. If more than one wear mechanism can be identified in a tribosurface, an inexperienced person examining the surface tends to overestimate the importance of large, spectacular scars, even if they are sparsely scattered over an otherwise smoothly worn surface. However, for cases where material loss is the lifetime-limiting damage, reducing the overall wear rate is the primary task. The wear rate, in turn, is determined by the volume of the produced wear particles and the frequency at which they are produced. If large scars cover only a small portion of the surface, they are the result of rare events, even more so if they are deep. In fact, for steady-state wear, the relative amount by which a mechanism contributes to the total wear is approximately proportional to the area fraction occupied by the scars of that mechanism. Thus, in combating the wear of devices with rival mechanisms operating, efforts should be focused on the mechanism that damages the largest surface area fraction and not on the mechanism that produces the largest individual scars. An example of surface damage caused by two competing wear mechanisms on the thread of a percussion drill coupling is shown in Fig. 3.

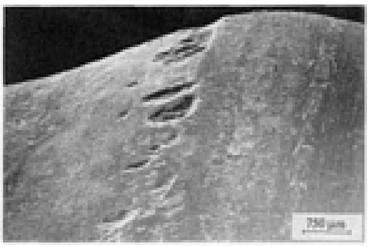


Fig. 3 Scanning electron micrograph of a thread of a percussion drill coupling. Mild abrasive and coarse pitting wear mechanisms dominate the wear to a fraction of 85% and 15%, respectively.

Selected Tribographical Examples

A number of tribographical examples of surface damage will be discussed in this section, applying the damage classification system shown in Fig. 1. The examples range from quite idealized to complex combinations of damage types.

Surface Damage Without Exchange of Material

Damage by microstructural changes, practically free from accompanying plastic deformation, is observed in ball bearing rings after many revolutions at high load. At a depth under the raceway, corresponding to maximum shear stress, the steel undergoes a phase transformation from tempered martensite to a dark etching zone of fine-grained high-carbon ferrite and residual martensite. At the center of the zone, a second transformation generates two differently oriented families of ferrite disks (Fig. 4). The excess carbon from the martensite is redistributed to form carbide lamellae between the ferritic disks. The driving force for these transformations is provided by the steep elastic stress gradients associated with each ball passage, which promote carbon diffusion in the martensitic matrix.

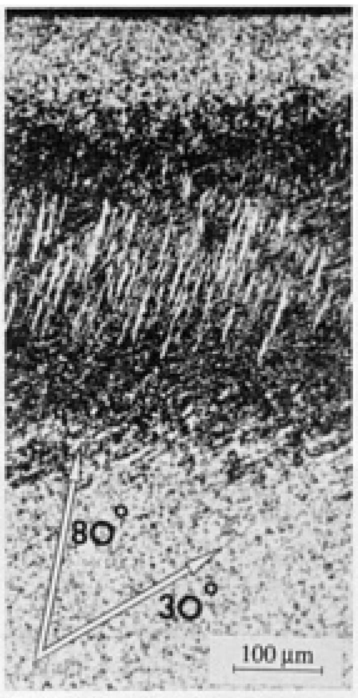


Fig. 4 Optical micrograph showing stress-induced phase transformation and successive development of a dark etching zone with parallel white etching ferrite disks under the raceway of an AISI 52100 steel ball bearing ring

Similar stress-induced changes in martensitic structure are frequently observed beneath the contact surfaces of gear teeth. It is not yet fully understood whether this kind of material transformation is beneficial or detrimental to the load-carrying capacity and wear resistance of ball bearings, gears, and similar components. It has been observed, however, that continued running at high load will eventually cause spalling, with associated cracks propagating preferentially along the new phases.

Damage Dominated by Plastic Deformation. Overloaded tribological components can suffer from gross plastic deformation. An example of this is the pearlitic carbon steel in Fig. 5. Extreme conditions of stress and temperature close to the surface soften the material locally, thus making it prone to plastic deformation. The outermost surface layer has reached a temperature above the austenitization level, which is evident from the fact that the relaxed layer had a microhardness of about 1200 HV, indicative of a later martensitic transformation (see below).

Fig. 5 Optical micrograph of a pearlitic carbon steel sample plastically deformed and phase transformed by a single-pass sliding contact against a cemented carbide cylinder. A continuous layer of friction martensite has formed in the superficial layer, while a dual phase of pearlitic and ferritic grains is found in an intermediate layer. Toward the surface, the pearlite grains have been partially transformed to martensite and the ferritic grains have recrystallized. No material has been lost from the surface.

The example shown in Fig. 5 was collected from sliding experiments against a cylindrical cemented carbide countersurface, with the cylinder axis perpendicular to the direction of sliding. Because of a vanishingly small angle of

attack between the contacting surfaces, there was virtually no loss of material. This form of damage is similar to that intentionally produced by the burnishing surface modification technique.

This example also illustrates the previous discussion about active and relaxed layers. Originally, the material had a homogeneous pearlitic structure. During the triboevent, the surface layer was heavily deformed and the temperature increased enough to transform the top layer into austenite, which consequently was its active phase. Transformation of austenite to untempered martensite occurred immediately after the triboevent by quenching through heat dissipation into the bulk material.

The importance of employing cross-sectional micrographs to fully understand the response of a material to tribological contacts is evident. The fact that the structure, and thus the mechanical properties of the surface layer, have been totally transformed cannot be ascertained from topographical imaging alone.

Plastic deformation of a surface without loss of material may also occur under the action of abrasive grains of nonaggressive geometry. Grooves surrounded by ridges are formed by local plastic flow of material. Such deformation may be detrimental when a high surface finish is required. Similarly, plastic deformation without loss of material often occurs in the initial phase of particle erosion of ductile materials (Fig. 6).

Fig. 6 Scanning electron micrograph showing plastic deformation without material loss in a cemented carbide surface exposed to single-particle impact. The binder phase (cobalt) has been extensively plastically deformed, while the hard carbides have been pushed around without deforming but have occasionally partially deadhered from the matrix. The surface has also been damaged by the pickup of fragments from the erosive SiC particle. Courtesy of Mikael Olsson

Damage Dominated by Surface Cracking. Surface cracks in tribosurfaces are usually caused by cyclic straining, which in turn is generated by some combination of cyclically varying frictional forces and frictional heating. The crack nucleation requires the presence of tensile stresses and a relatively brittle surface material. In principle, there are two ways in which tensile stresses can be generated in a tribocontact:

- Heat expansion of the surface layer beyond the compressive yield limit, followed by cooling, introduces tensile stresses in the plastically deformed layer
- Friction force directly generates tensile stresses at the trailing end of tribocontacts (Fig. 2)

In addition, materials which initially are ductile may become progressively embrittled after cyclic straining, and thereby prone to crack nucleation. The main plane of crack propagation is normal to the direction of the tensile stress.

Excessive strains caused directly by the friction force usually result in a pattern of parallel cracks perpendicular to the sliding direction, whereas friction-induced thermal cycling yields isotropic strains and a corresponding network pattern (Fig. 7). Because the cracks do not primarily involve material losses, they may well be tribologically harmless, unless they lead to secondary effects, such as enhanced corrosive attack (stress corrosion), reduced fatigue life, or spalling. Tribocracks are less likely to be observed in systems with ductile materials. At high wear rates, where the outermost layer is continuously worn off at a rate higher than that at which crack propagation would have occurred in the absence of wear, tribocracking is also suppressed.

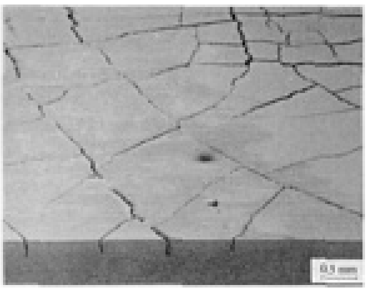


Fig. 7 Scanning electron micrograph of surface and cross sections showing cracks in a hydraulic steel piston. The reciprocal sliding was horizontal.

Surface Damage Involving Loss of Material: Wear

Damage Dominated by Shear Fracture. Plastic shear deformation of a surface is one of the telltale signs of adhesive wear. Depending on the application and the severity of the deformation, smearing, galling, scuffing, scoring, or seizure are also signs of adhesive wear. Typical of severe shear fracture is the generation of a scalelike topography (Fig. 8). Occasionally, fragments lost from the mating surface can be picked up and observed at the end of the groove at which they have just formed (see "A" in Fig. 8). Another groove, formed in a similar manner, can be seen at the top left of Fig. 8. Note that a picked-up fragment may be capable of scoring its own material if it has been strengthened during previous triboevents by successive mechanical and thermal cycling, causing a gradual decrease in grain size and, in the case of steels, repeated martensitic transformations (Fig. 5).

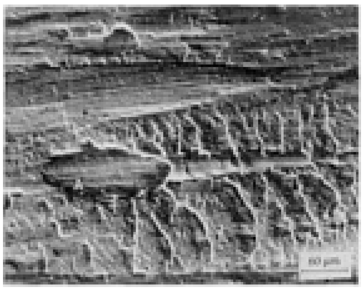


Fig. 8 Scanning electron micrograph showing damage dominated by shear fracture typical of severe adhesive wear. Pin surface from pin-on-disk sliding between two identical AISI D1 steel specimens. Sliding direction of counterface was from right to left.

This type of severe surface damage is indicative of a system that in action showed high friction, high wear rate, and a high noise level. It usually terminates the life of sliding components and should by all means be avoided, at least by selection of different materials for the two mating surfaces if not by redesign or improved lubrication.

Another instructive example of complex surface damage is presented in Fig. 9, which shows the microstructure of the subsurface layer of a special brass component that has been slid against steel. The grain morphology has been affected by massive plastic flow in the direction of counterface sliding, changing from the arbitrary shapes at great depth, via elongation into thin parallel disks, to a hardly resolvable microstructure of very fine grains close to the surface. The superficial layer has also undergone dynamic recrystallization during continued sliding and frictional heating.

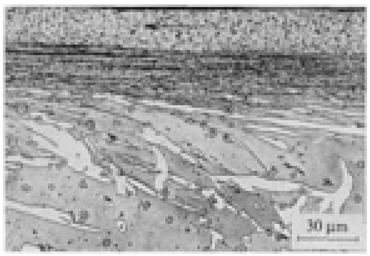


Fig. 9 Optical micrograph showing the subsurface morphology of special brass damaged by shear fracture, plastic deformation, and microstructural changes, when slid against steel

As in the case of the steel surface shown in Fig. 8, the mechanism of material loss for the example illustrated in Fig. 9 was severe shear fracture. Initially, brass patches adhered to the steel surface; loose, flakelike brass particles were subsequently produced at an unacceptably high rate.

Damage by shear fracture is directly related to the shear force transmitted between the mating surfaces. This force, in turn, is related to the tendency of different mating materials to form solid solutions or intermetallic compounds with each other. In general, materials having a hexagonal close-packed (hcp) structure exhibit lower adhesive wear than either a body-centered cubic (bcc) or a face-centered cubic (fcc) structure material. Additional information is available in the article "Sliding and Adhesive Wear" in this Volume.

Damage Dominated by Extrusion. Extrusion is a type of surface damage that can occur in certain tribological systems with sliding or rolling contact. Material is plastically squeezed ahead of a traversing asperity or a rolling element. The latter case is exemplified by Fig. 10, which shows extrusion of surface material from the originally hard chromium-plated surface of a carbon steel loop. The loop was used to attach a steel spring to the frame of a hospital bed. The rocking motion of the spring against the loop broke up the chrome layer and squeezed the underlying steel to the sides of the contact area.

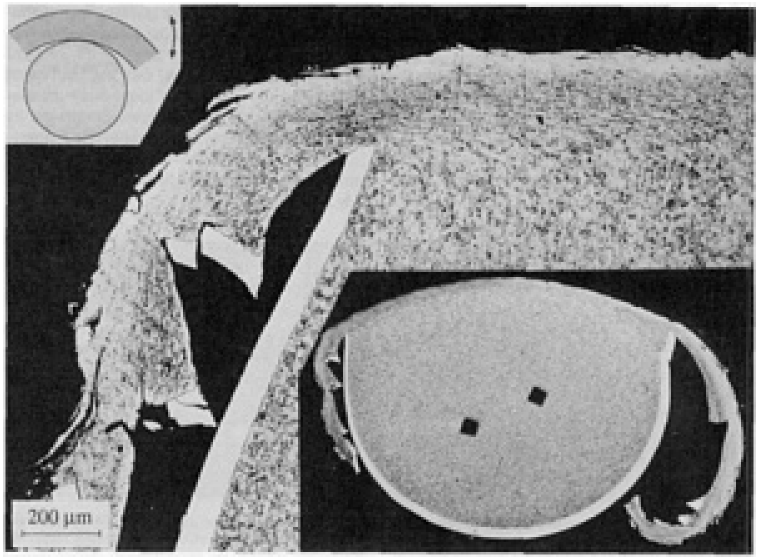


Fig. 10 Optical micrograph showing extrusion of surface material from a hard chromium-plated steel loop. Insets: overview and contact geometry

The extrusion of burrs limited the life of the loops both by gradually decreasing their strength and by generating a potential source of injury for the hospital staff. About 50 loops in each of about 5000 beds had to be replaced. The key to avoiding these premature failures was using a harder steel in the loops.

Damage Dominated by Chip Formation. The plowing action of a hard abrasive particle against a softer surface may cause plastic deformation without material loss. Conditions can also be such that a chip is cut out of the surface (Fig. 11). After the chip is lost, the surface scar does not primarily reveal the loss of material, but the abrasive wear can generally be identified by, for instance, the associated cumulative changes of the bulk shape of the worn part. Material loss by chip formation is usually accompanied by heavy plastic deformation of material in the ridges and under the grooves and further by microstructural changes in a surface layer similar to those presented in Fig. 5 and 9. Typical surface damage resulting from abrasive wear of a road grader blade is shown in Fig. 12.

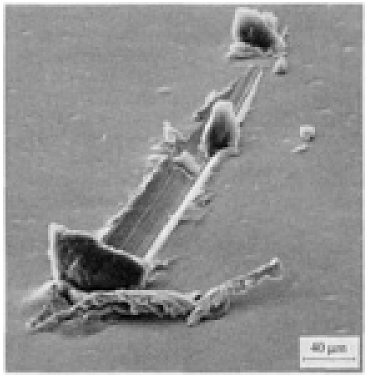


Fig. 11 Scanning electron micrograph showing surface damage by chip formation, plastic deformation, and pickup of fragments of a ceramic particle abrading a copper surface

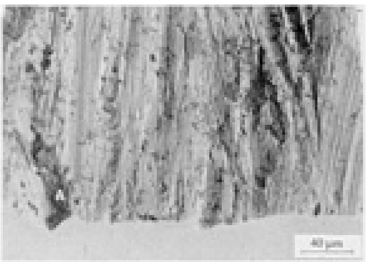


Fig. 12 Scanning electron micrograph of the surface and cross section of a road grader blade worn against gravel. The surface has been damaged by chip formation, plastic deformation, and pickup of silica, revealed as dark, rough patches at A.

Damage Dominated by Tearing. Tearing is a mode of material separation typical of polymeric materials. It is initiated by local stiction between the mating surfaces, followed by material being pulled or torn out of the surface, often leaving elongated tongues or strings as wear debris or still loosely attached to the surface, as in Fig. 13. Figure 13 shows a small piston pump used in analytical equipment for medical applications. The prime cause of failure was leakage caused by local tearing, which destroyed the surface topography. Furthermore, wear particles contaminated the pumped medium.

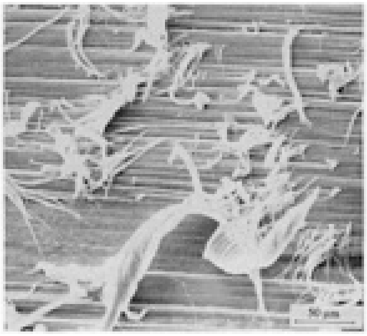


Fig. 13 Scanning electron micrograph showing damage caused by tearing in the cylinder surface of a small polytetrafluoroethylene (PTFE) piston pump used to portion human body liquids into an analytical instrument.

The tearing mechanism requires strong adhesion to the countersurface combine with an anisotropic cohesion of the material. Polymeric materials often fulfill the latter condition by having a high degree of structural texture.

The material being torn out of the surface may also form rolls (Fig. 14). This mechanism of material loss is best characterized by rubber erasers.

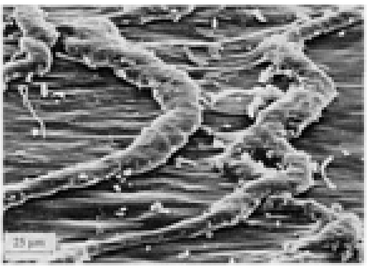


Fig. 14 Scanning electron micrograph showing roll formation associated with tearing of material from the surface of polyurethane during sliding against a flat metal surface. Sliding direction of counterface was from left to right.

Damaged Dominated by Brittle Fracture. Brittle fracture is a common mode of surface failure of hard and relatively brittle materials such as ceramics, ceramic coatings, cemented carbides, cermets, and hardened steels. It is initiated by tensile stresses high enough to exceed the fracture stress of the surface material in its active state. These stresses can be caused by a number of tribomechanisms. The type of damage can vary from a relatively large-scale fracture (Fig. 15) to fragmentation on a microscale (Fig. 16). It is obvious that the tool shown in Fig. 15 has lost its original shape and thus its performance. The erosion rate of the alumina shown in Fig. 16 was unacceptably high in its intended application as a mechanical face seal, but could be dramatically reduced by toughening the alumina matrix with SiC whiskers. Scratching by abrasive particles can also cause brittle fracture. In the case of optical lenses and protective glass surfaces, this type of damage sets the life limit (Fig. 17).

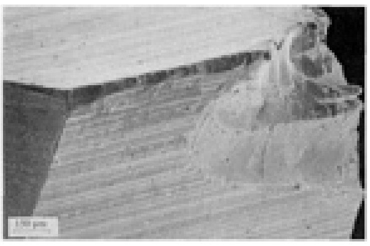


Fig. 15 Scanning electron micrograph showing brittle fracture of a cemented carbide tool tip used to cut aluminum

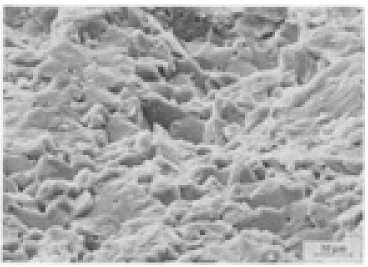


Fig. 16 Scanning electron micrograph showing brittle fracture by particle erosion of an Al_2O_3 surface (250 μm SiC particles, 70 m/s velocity, 90° angle of impingement). Courtesy of Mikael Olsson

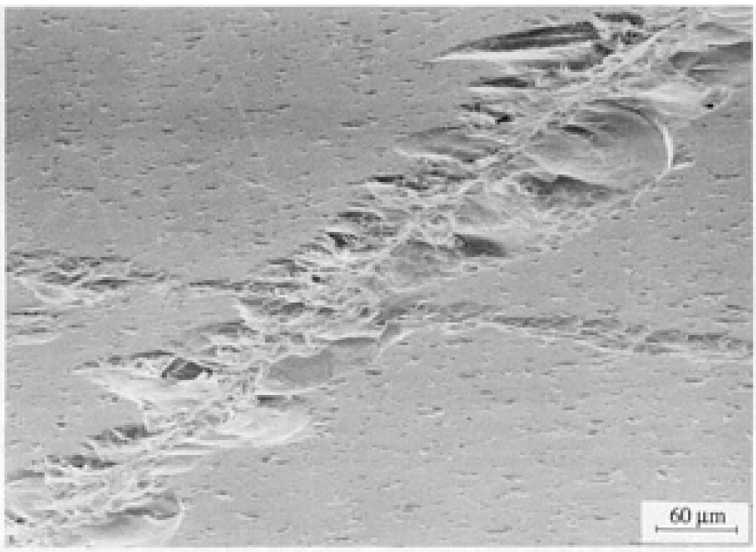


Fig. 17 Scanning electron micrograph showing brittle fracture in a watch glass caused by scratching against hard particles. The scratches impair the transparency of the glass.

Brittle fracture is often associated with very high wear rates. To prevent this type of damage, high point loads should be avoided. In sliding contacts, this can be accomplished by using very smooth surfaces and rounded shapes and by avoiding large abrasive particles or sharp countersurfaces. Ceramic surfaces that have been damaged by brittle fracture develop a very rough and vitreous topography. Also, once the tribological response of brittle materials has changed from a mild wear regime to severe brittle fracture, the mild regime is most unlikely to be reestablished even if the load is strongly reduced. This is in contrast to ductile metals, where a severely worn surface often can be "healed" by continued mild sliding.

Damage Dominated by Fatigue Fracture. Surface fatigue fracture is usually the result of repeated unidirectional or reciprocal sliding and is often seen on the same types of materials that suffer brittle fracture. In addition to hardness and toughness in the active state, the thermal expansion coefficient of the surface material plays a major role.

Figure 18 shows a punch tool used to cut austenitic stainless steel that reveals a characteristic crack pattern perpendicular to the direction of work material sliding. Thus, the cracks are caused primarily by the friction force, and not by thermal cycling. Initially, no material loss was associated with the cracking, and consequently tool performance did not suffer. However, spalling began to accelerate the wear after a certain number of cuts, thereby increasing the punch clearance and deteriorating the tolerances of the cut product. Finally, one of the transverse cracks initiated a large-scale fracture of the punch edge, instantly ending the tool life. Figure 19 illustrates another example of surface fatigue fracture, combined with plastic deformation of the surface to a mirror finish.

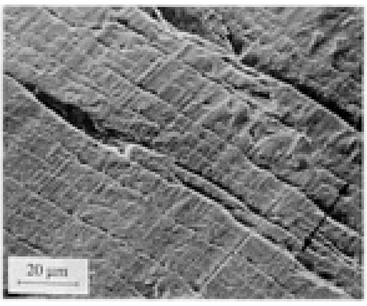


Fig. 18 Scanning electron micrograph showing fatigue fracture in a tool steel (AISI D1) punch used to cut 1 mm (0.04 in.) stainless steel sheets

Fig. 19 Scanning electron micrograph revealing two damage mechanisms on a steel cam follower, run against a steel cam. Smoothing by plastic deformation and wear by surface fatigue-induced fracture. Note the surface cracks partially surrounding the wear craters.

Low-amplitude oscillatory sliding between tribosurfaces, or so-called fretting, may lead to surface fatigue fracture and fretting-accelerated corrosion when the amplitude is so small that mating contact surface protrusions reestablish contact during each half-cycle. Because of the convex shape of the protrusions, the contact pressure is highest at the center of the contact area and falls to zero at the rim. This leads to mixed stick-slip conditions, with a central patch where the surfaces adhere ("stick") and a surrounding annular area where slip occurs. The stick-slip transition region is characterized by high stress concentrations and is the site for surface fatigue crack nucleation (Fig. 20).

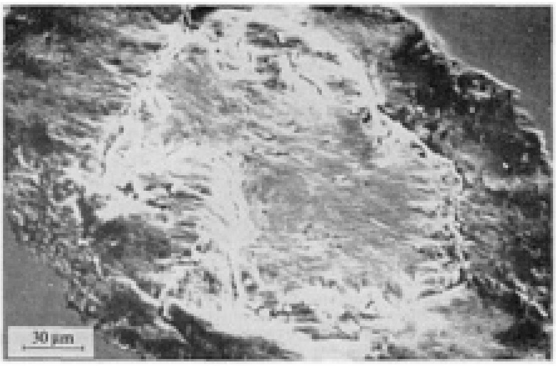


Fig. 20 Scanning electron micrograph showing surface damage in stainless steel, typical of fretting-induced fatigue fracture. A crack loop surrounds the unaffected stick patch.

Damage Dominated by Dissolution or Diffusion. A chemically aggressive environment may attack tribosurfaces and remove material by dissolution. In lubricated systems with chemically active lubricants, dissolution of the surface material may, in fact, dominate the material loss. The effect is promoted by frictional heating.

In metal cutting, the tool material has been observed to dissolve by solid-state diffusion through the tool/workpiece and tool/chip interfaces. This dissolution or diffusion mechanism can account for a considerable portion of the tool wear, particularly in wear-resistant cemented carbide tools used at high cutting speeds. It often acts preferentially--for example, for cemented carbides on the cobalt matrix.

It is generally difficult to diagnose damage by dissolution or diffusion, because it does not leave any signs in the form of scars or phase transformations. This mechanism should be considered if the wear rate is very low for the prevailing contact temperature and the topography is extremely smooth on single-phase materials or if chemically more inert phases of multiphase materials protrude from an otherwise smooth surface.

Damage by Interaction Between Corrosion and Fracture. A typical example of the appearance of surfaces damaged by corrosive wear is illustrated in Fig. 21, which shows two sections of links of a galvanized steel chain used to anchor a small dingy. In service, the chain was frequently subjected to light stretching and some movement, which led to rubbing between individual links. Links that were above the water surface suffered negligible corrosive attack and displayed negligible wear. The submerged links suffered only mild corrosion, but experienced very severe wear in those parts that were in mechanical contact. In this case, the cyclic removal of the parabolically growing oxide layer and the resulting exposure of the bare metallic surface to water generated the high wear rate. Above the water surface, the mechanical contact between the links was identical, but the corrosive conditions were less aggressive, and the zinc coating was not even penetrated.

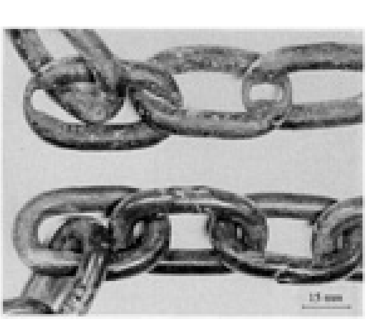


Fig. 21 Wear by corrosion and shear fracture of two sections of a galvanized steel chain partly submerged in water. Note the extreme difference in wear rate between the underwater (top) and above-water (bottom) links, caused by the synergistic effects between corrosion and wear. The general corrosion rate was low.

One possible mechanism of interaction between corrosion and material loss due to fracture is clearly illustrated by the particle erosion of an oxidized surface in Fig. 22. The oxide layer is spalled off in an area much larger than the impact crater, while the substrate is almost unaffected.

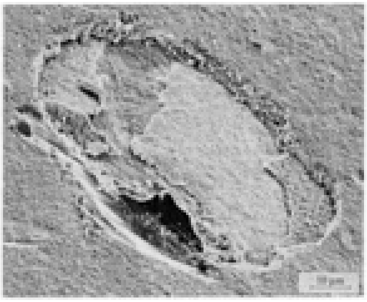


Fig. 22 Scanning electron micrograph showing particle-impact-induced brittle fracture that caused spalling off of oxide layers on cobalt. Courtesy of Asa Hammarsten

Surface Damage Involving Gain of Material

Transfer of material from one of the tribosurfaces to the other or pickup of loose wear debris occurs frequently in many tribological situations, particularly in the adhesive wear of ductile materials. Generally, the added material may:

- Clad to the surface (Fig. 23)
- Become embedded in the surface (Fig. 24)
- Successively build up a more or less complete coating on the surface (Fig. 25)

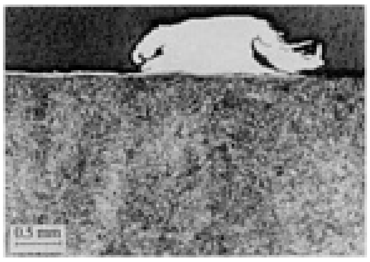


Fig. 23 Optical micrograph of cross section of a patch of wear debris on a quenched and tempered steel surface, transferred from the high-speed steel (HSS) counterface during sliding contact in a crossed-cylinders test. Having been in constant contact with the countersurface and thus frictionally heated to a higher temperature, the normally more wear-resistant HSS material has been worn by adhesive shear fracture, while the countersurface has been damaged by the transferred debris. Courtesy of Mikael Olsson.

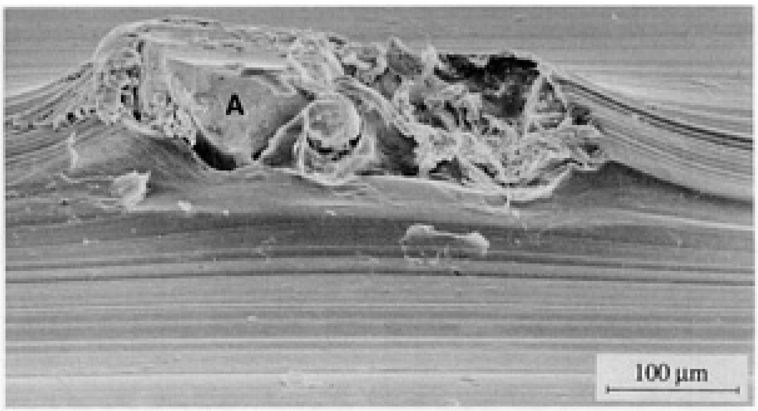


Fig. 24 Scanning electron micrograph showing metallic wear debris at A, picked up from the counterface material and embedded in a PTFE surface during reciprocating sliding. The particle acted highly abrasively on the countersurface and has been moved cyclically back and forth by the friction forces.

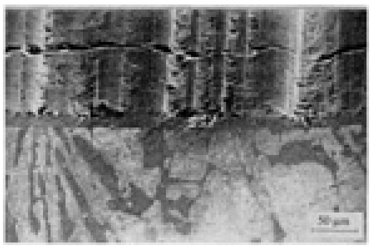


Fig. 25 Scanning electron micrograph showing the top surface and cross section of a hot roll with a protective oxide layer transferred from the heat (work material). As long as the layer adheres to the roll surface, the oxide coating protects the underlying metal efficiently by continuous pickup of oxide at the same rate as the simultaneous loss by wear. Courtesy of Lynn Erickson

Transferred material may be either detrimental or favorable to the performance of a tribosurface. In some cases it can protect the surface from wear, while simultaneously causing an increase in the wear rate of the countersurface, as demonstrated by Fig. 23 and 24. It can also efficiently protect the coated surface without harming the countersurface, as shown by Fig. 25. In this case, in addition to the high-temperature oxidation properties of the worked carbon steel, the composition and surface temperature of the roll material are crucial factors for the formation of the oxide layer. In service, as a matter of fact, rolling with the initial metallic roll surface should be avoided; surprisingly, a surface "damaged" by picked-up oxide is ideal for rolling.

Damage Dominated by Contact-Generated Corrosion. Tribo-accelerated corrosion is a common form of surface damage. The corrosion is often problematic as it leads to an increased wear rate or to a disadvantageous topography. In the case of fretting corrosion of electrical contacts (Fig. 26), the oxide layer can be life-limiting by building up an unacceptably high contact resistance. The corrosion is accelerated by vibrations that facilitate exchange of atoms across the oxide layer, both by forming microcracks and by directly promoting diffusion.

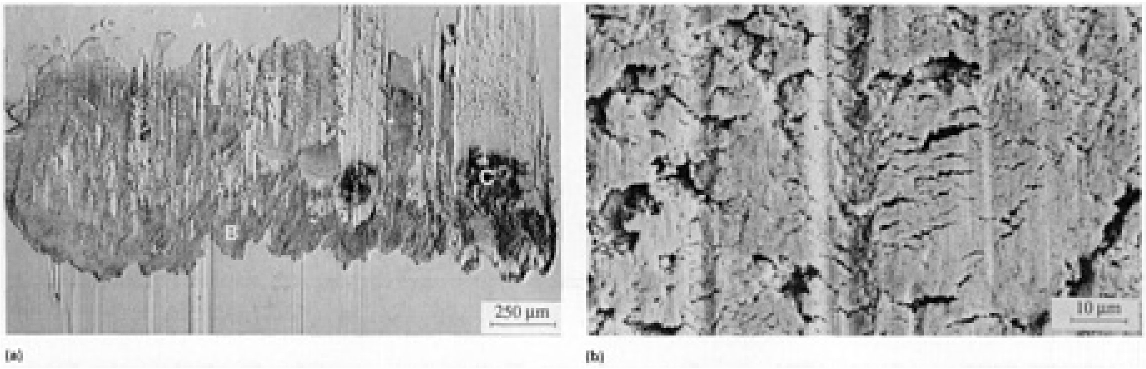


Fig. 26 Scanning electron micrograph showing surface damage by corrosion and plastic deformation of a tin-coated copper electrical contact subjected to fretting corrosion. Direction of vibration was vertical. (a) Overview: The light gray areas (A) are unaffected tin plating, the medium dark areas (B) are oxidized tin, and the darkest areas (C) are revealed copper surface. (b) Detail: The tin oxide layer was also damaged by extensive surface cracking. Courtesy of Åsa Kassman

Tribography as a Tool

The previous sections of this article have classified and illustrated the appearance of surface damage and have related it to the generating mechanisms and processes. The next step would be to formulate a "recipe" of actions to improve the tribological behavior of a system. This subject will not be covered here, but a few examples are offered to demonstrate a fruitful method of reasoning, utilizing the findings of a tribographical investigation.

Whenever a surface is shown to have suffered from *plastic deformation,* a number of counter-measures are possible. Plastic deformation is caused by exceeding the yield stress of the material in the active state. The applied stress can be reduced by increasing the load-bearing area, decreasing the load, applying more efficient lubricants, and so on. The yield stress can be increased by improving the cooling and thus reducing the softening in the active state, by exchanging the material, or by applying a hard surface coating.

Brittle surface fracture is initiated whenever the fracture strain of the material in its active or relaxed state is exceeded. In sliding contacts, this can be avoided by reducing the shear stresses by lubricating more efficiently, by promoting the formation of friction-reducing oxide layers, and so on. Furthermore, friction stresses can be reduced by redesigning for lower normal forces or increased load-bearing area.

When *surface cracks* are discovered by tribography, the pattern formed reveals details of the contact situation. Formation of more or less parallel cracks is indicative of excessive shear strains in the direction perpendicular to the cracks. Network patterns are indicative of isotropic strains, usually induced by cyclic thermal expansion.

Abrasive Wear

Joseph H. Tylczak, Albany Oregon

Introduction

ABRASIVE WEAR, as defined by ASTM (Ref 1), is due to hard particles or hard protuberances that are forced against and move along a solid surface. Wear, in turn, is defined as damage to a solid surface that generally involves progressive loss of material and is due to relative motion between that surface and a contacting substance or substances.

The cost of abrasion is high and has been estimated as ranging from 1 to 4% of the gross national product of an industrialized nation. The effect of abrasion is particularly evident in the industrial areas of agriculture, mining, mineral processing, earth moving, and essentially wherever dirt, rock, and minerals are handled. Examples include plows, ore loading/moving buckets, crushers, and dump truck beds.

When two surfaces contact, wear occurs on both surfaces. Individuals and industry tend to focus on th wearing surface that has the greatest potential for their own economic loss, and consider the other surface to be the abrasive. For example, an individual walking up the stairs of a building would be more likely to think that his shoes, rather than the stairs, were experiencing abrasive wear, whereas the maintenance staff would have the opposite opinion. In actually, both surfaces are being subjected to abrasive wear.

The rate at which the surfaces abrade depends on the characteristics of each surface, the presence of abrasives between the first and second surfaces, the speed of contact, and other environmental conditions. In short, loss rates are not inherent to a material. With reference to the above example, changing the material of either the shoes or the steps could, and often would, change the wear on the opposite counter-face. The addition of an abrasive, such as a layer of sand, on the steps would further change the situation, in that the sand would be the second surface that contacts both the shoes and the steps.

Abrasion is typically categorized according to types of contact, as well as contact environment. Types of contact include two-body and three-body wear. The former occurs when an abrasive slides along a surface , and the latter, when an abrasive is caught between one surface and another. Two-body systems typically experience from 10 to 1000 times as much loss as three-body systems for a given load and path length of wear. Contact environments (Fig. 1) are classified as either open (free) or closed (constrained).

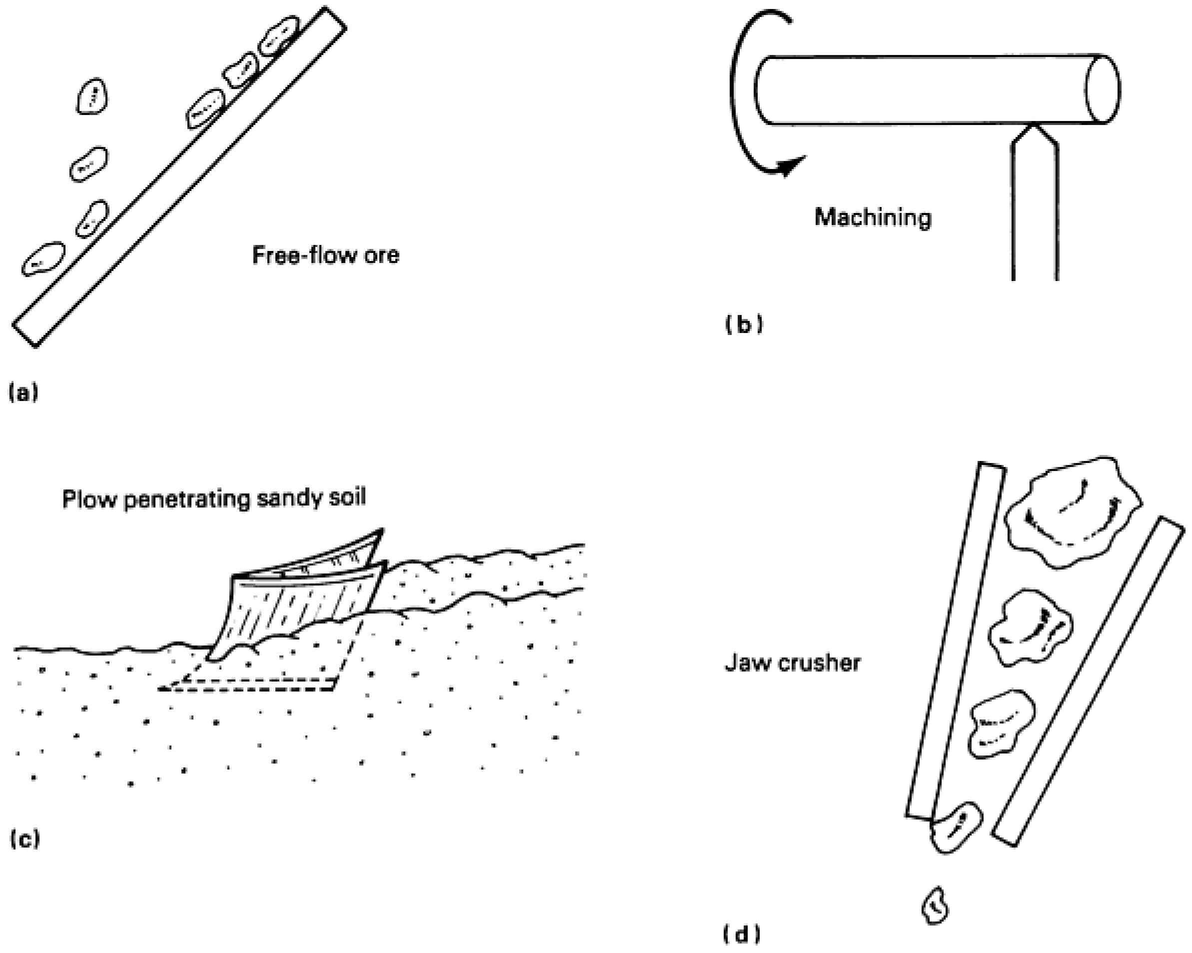


Fig. 1 Types of contact during abrasive wear. (a) Open two-body. (b) Closed two-body. (c) Open three-body. (d) Closed three-body

In several different tests, Blickensderfer *et al.* (Ref 2) showed that for a given load and path length of wear, the wear rate is about the same for both open and closed systems. However, measurements of the loss in closed systems will often appear higher than the loss in open systems. This probably occurs because most closed systems experience higher loads.

Abrasion is often further categorized as being low stress, high stress, or gouging. Low-stress abrasion occurs when the abrasive remains relatively intact, for example, when sanding wood with sandpaper. High-stress abrasion exists when abrasive particles are being crushed, for example, in a ball mill where both the grinding balls and the ore are down. In gouging abrasion, a relatively large abrasive will cut the material that is not fully work hardened by the process from the material of concern, for example, when rocks are crushed in jaw crusher.

Several mechanisms have been proposed to explain how material is removed from a surface during abrasion. These mechanisms include fracture, fatigue, and melting. Because of the complexity of abrasion, no one mechanism completely accounts for all the loss. Figure 2 depicts some of the processes that are possible when a single abrasive tip traverses a surface. They include plowing, wedge formation, cutting, microfatigue, and microcracking.

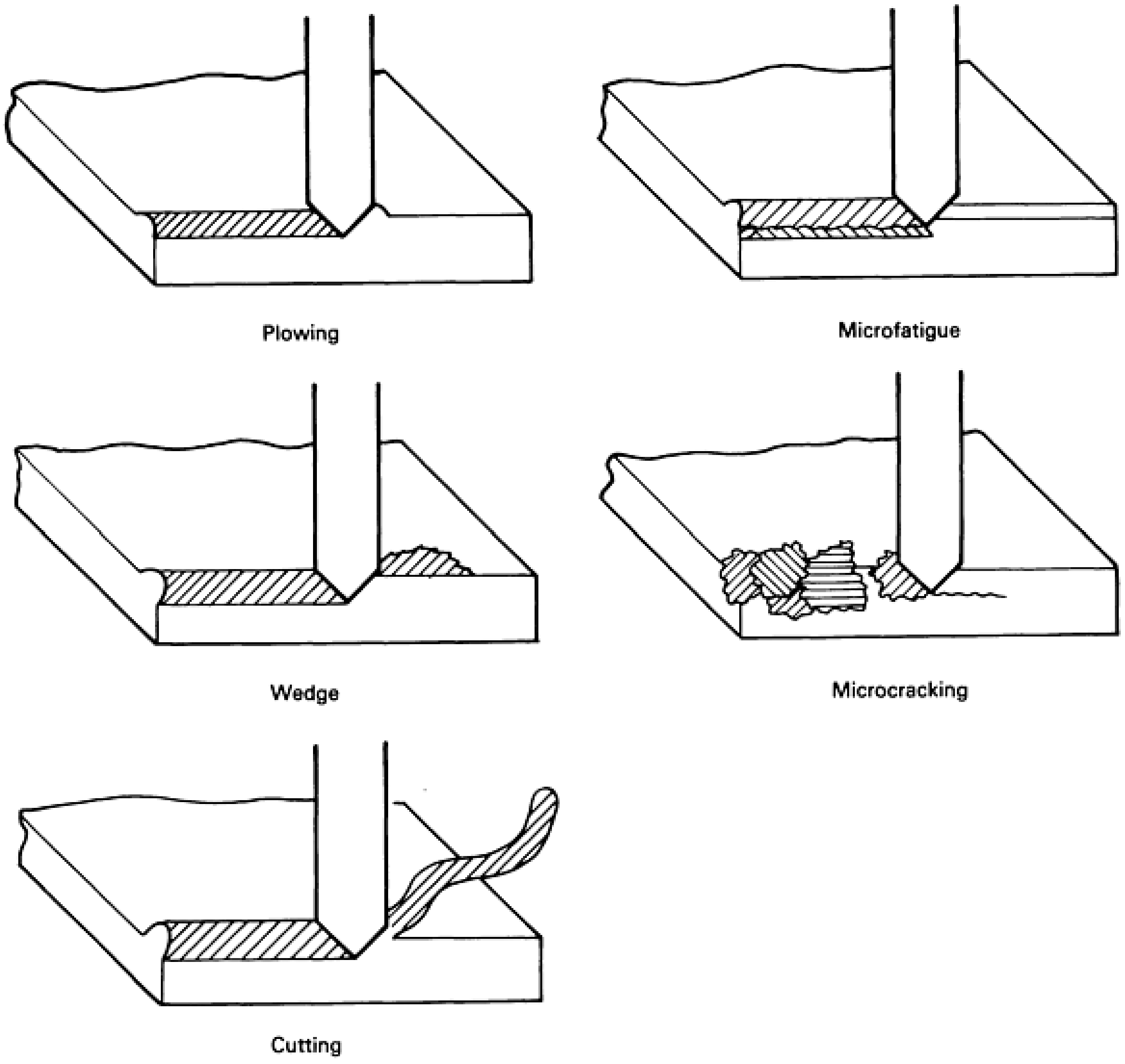


Fig. 2 Five processes of abrasive wear

Plowing is the process of displacing material from a groove to the sides. This occurs under light loads and does not result in any real material loss. Damage occurs to the near surface of the material in the form of a build up of dislocations through cold work. If later scratches occur on this cold-worked surface, then the additional work could result in loss through microfatigue.

When the ratio of shear strength of the contact interface relative to the shear strength of the bulk rises to a high enough level (from 0.5 to 1.0), it has been found that a wedge can develop on the front of an abrasive tip. In this case, the total amount of material displaced from the groove is greater than the material displaced to the sides. This wedge formation is still a fairly mild form of abrasive wear.

The most severe form of wear for ductile material is cutting. During the cutting process, the abrasive tip removes a chip, much like a machine tool does. This results in removed material, but very little displaced material relative to the size of the groove. For a sharp abrasive particle, a critical angle exists, for which there is a transition from plowing to cutting. This angle depends on the material being abraded. Examples of critical angles range from 45° for copper to 85° for aluminum (Ref 3, 4). Abrasion is not dependent on scratches by carefully oriented abrasive grains. Kato (Ref 5) and others have analyzed the effect of a rounded tip pushing through a surface.

For ductile materials, the mechanisms of plowing, wedge formation, and cutting have been observed (Fig. 3). It was found that the degree of penetration was critical to the transition from plowing and wedge formation to cutting. When the degree

of penetration, defined as depth of penetration divided by the contact area, exceeded about 0.2, cutting was the predominant mode of wear.

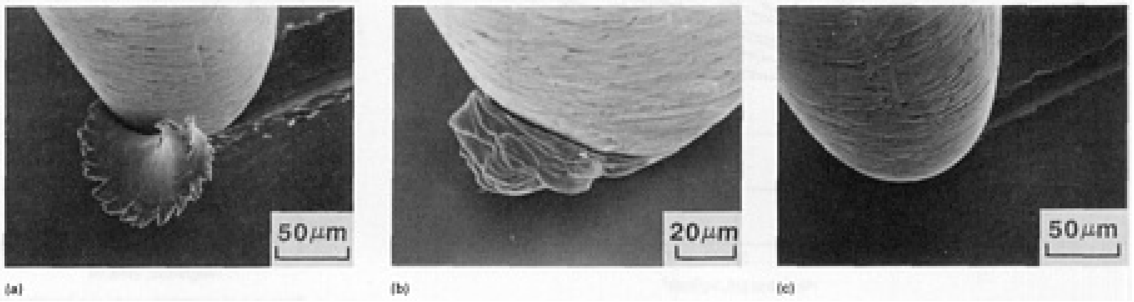


Fig. 3 Examples of three process of abrasive wear, observed using a scanning electron microscope. (a) Cutting. (b) Wedge formation. (c) Plowing. Source: Ref 5

When an abrasive grain abrades while cutting a surface, the maximum volume of wear that can occur is described by:

$$W = Ad \qquad \textbf{(Eq 1)}$$

where W is the volume of material removed, A is the cross-sectional area of the groove, and d is the distance slid. The cross-sectional area of the groove A is dependent on the abrasive grain shape and the depth of penetration, p:

$$A = k_1 p \qquad \textbf{(Eq 2)}$$

where k_1 is constant-dependent on the shape. In turn, the depth of penetration, p, is again dependent on the shape of the grain; the load, L; and the hardness, H, of the material:

$$p = k_2 \frac{L}{H} \qquad \textbf{(Eq 3)}$$

Many factors affect k_1: the possibility of plowing rather than cutting; the abrasive grain may roll and avoid wear; the abrasive grain may break down and not be effective during the latter part of its contact path; and others. Equation 1, 2, and 3 can be combined, forming:

$$W = k_3 \frac{Ld}{H} \qquad \textbf{(Eq 4)}$$

This is commonly known as Archard's equation (Ref 6), which was derived for adhesive wear but has proven very useful in abrasive wear, as well. Factors affecting k_3 are addressed below.

Commonly, materials are described as having good or bad wear resistance, R, which is simply defined as the reciprocal of wear volume:

$$R = \frac{1}{W} \qquad \textbf{(Eq 5)}$$

Brittle materials have an additional mode of abrasive wear, namely, microfracture. This occurs when forces applied by the abrasive grain exceed the fracture toughness of the material. This is often the predominant mode of severe wear for the ceramic materials, and is active in materials such as white cast irons.

Melting or other thermally related mechanisms of loss are also possible. The melting theory depends on small localized areas of strain-induced adiabatic heating. It has been shown that there does seem to be enough temperature rise for recovery to take place near the abraded surface (Ref 7).

Effects of Material Properties on Abrasive Wear

A variety of material characteristics have been shown to either form a correlation with abrasive wear or have some effect on it. These properties include hardness, elastic modulus, yield strength, melting temperature, crystal structure, micro-structure, and composition.

It has been shown experimentally and theoretically that the hardness of a material correlates with its abrasion rate. Khrushchov (Ref 8) performed a large amount of testing and found an inverse relationship between abrasion rate and annealed hardness for pure materials (Fig. 4). He also tested steels of varying hardness. The hardnesses were inverse linearly related to abrasive wear, except that they had a different slope from that of the pure materials.

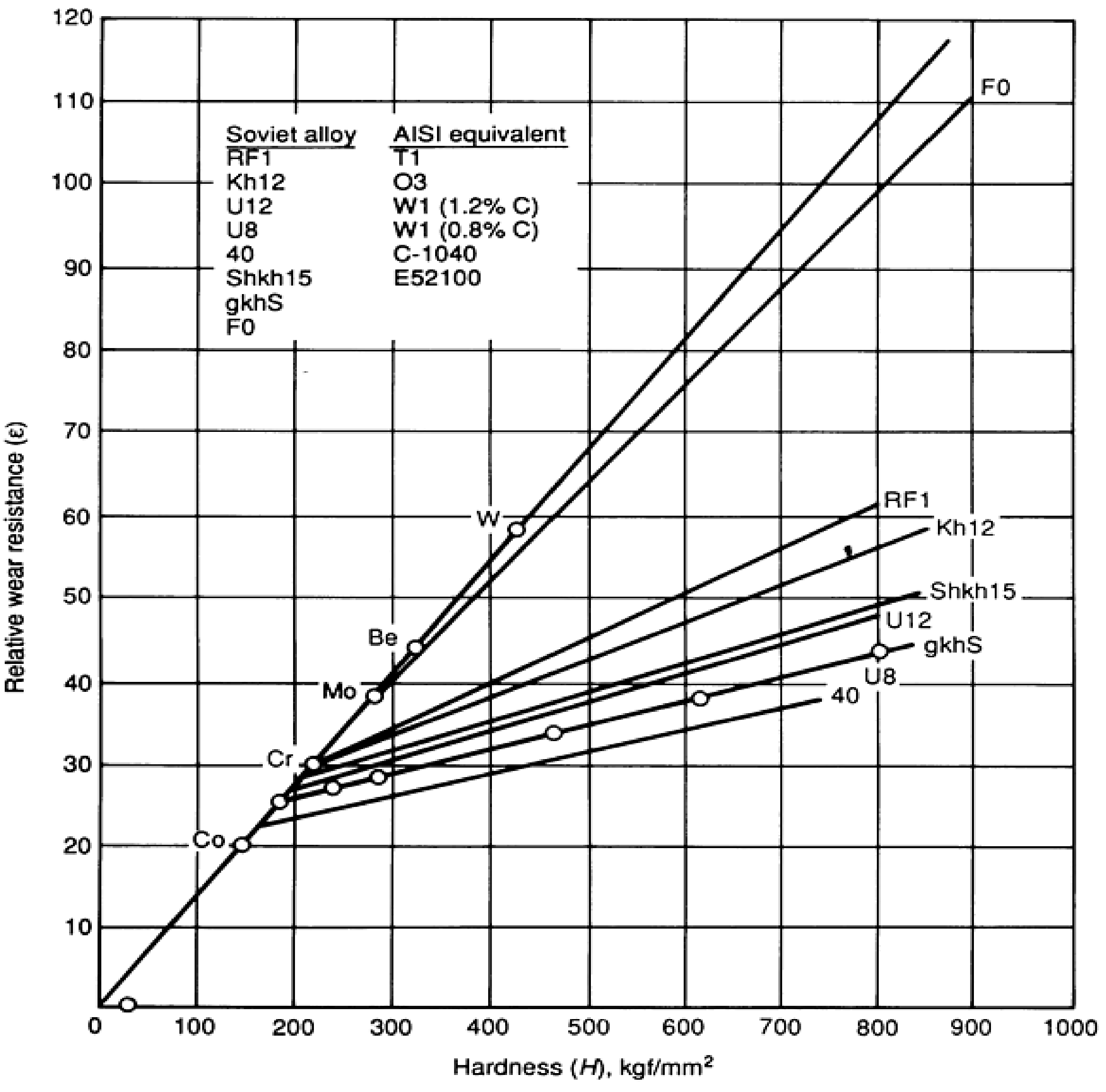


Fig. 4 Wear resistance versus hardness for pure metals and alloys. Source: Ref 8

It is generally thought that the surface of a material is work hardened up to a very high level during the process of abrasion. Richardson (Ref 9) investigated work hardening by plowing during wear on a group of pure metals and steels. He compared the resulting hardness of the surface to surfaces hardened by shot blasting and trepanning, and found that

abrasion produced a high hardness that was nearly the hardness of trepanning. In addition, the wear resistance of the metal was proportional to the hardness of the worn surface.

Abrasive wear has also been found to be dependent on crystal structure and orientation. Alison (Ref 10) showed that cubic metals wear at about twice the rate of hexagonal metals, which was attributed to the lower work-hardening rate of the hexagonal metals. In addition, Steijn (Ref 11) studied the wear of single crystals. Scratching body-centered cubic (bcc) and face-centered cubic (fcc) metals with a prepared surface on the (001) plane, he showed wider scratch width, which implied higher wear, along the <100> than the <110>direction.

Microstructure is also important. Austenite and bainite of equal hardness are more abrasion resistant than ferrite, pearlite, or martensite. This is because of the higher strain-hardening capacity and ductility of austenite.

Additionally, it has been found that fracture toughness, K_{Ic}, of the material is important in determining abrasive wear for ceramics and, to a lesser degree, white cast irons. Fischer (Ref 12) prepared a series of zirconia samples with constant hardness, but varying toughness. He found that the wear decreased with the fourth power of the toughness (Fig. 5). This fourth power law applies to a single case of material and test parameters, but it does show the important effect of toughness on brittle materials.

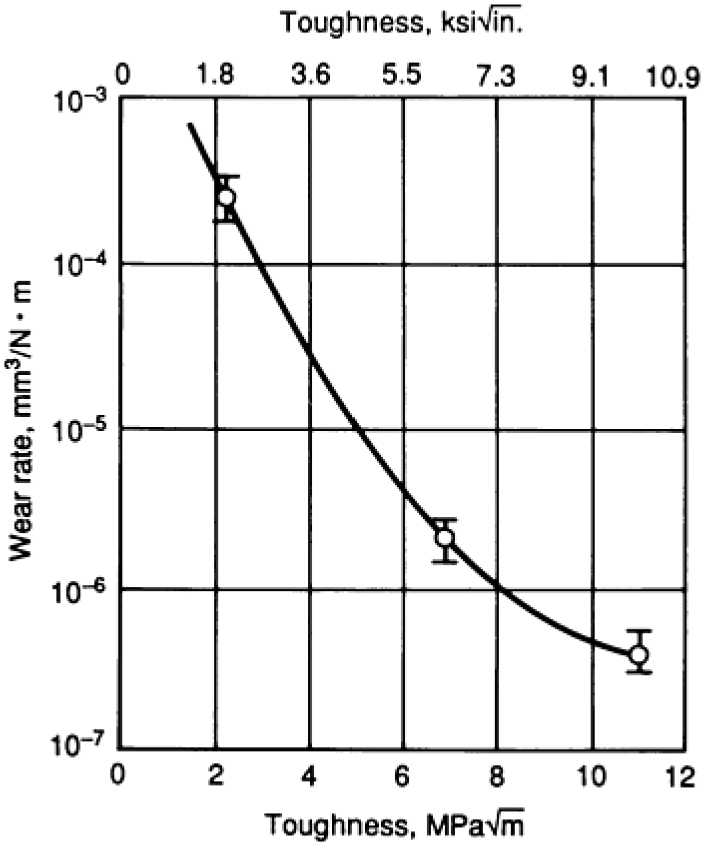


Fig. 5 Wear rate of zirconium oxide as a function of fracture toughness. Source: Ref 12

Alloying is often used to improved the performance of a material. These additions can take either interstitial or substitutional locations. Adding carbon to iron is a good example of an interstitial addition used to improve abrasion resistance. Tylczak (Ref 13) studied the abrasion resistance of Zr and Ti alloys with small interstitial additions of N or O. Like carbon in iron, these alloys also decreases of wear with small increases of interstitial content. For substitutional alloy systems, he also showed that the abrasion of alloys with complete solid solubility, such as Hf-Zr, Cu-Ni, and Cr-V, follows a law of mixing, where the abrasion is proportionate to the amount of each alloy. Abrasion was also found to be somewhat affected by solidus temperature and hardness. For solid-solution mixtures, this indicates that deviations from a law of mixing are separately dependent on strength of the bond and distortion of the crystal lattice.

A common way to modify the properties of a material is to produce a second phase. Treatments that cause the formation of precipitates can result in larger increases in hardness and yield stress. On the basis of the previously developed simple wear model, one would expect this to lead to large decreases in abrasive wear. Unfortunately, the small coherent particles are often sheared during plastic deformation, and the incoherent particles fail to block the dislocations that are generated. As a result, precipitation treatments are not generally a useful way to decrease abrasive wear.

Larger, hard incoherent precipitates or particles such as carbides can be useful in decreasing abrasive wear. Figure 6 shows what happens when the size ratio of the abrasive grains to the hard particles in the matrix is varied. When the incoherent particles are somewhat larger than the abrasive grains abrading the surface, they are generally effective in decreasing the total material wear. Examinations of worn surfaces have revealed that the abrasive grains are lifted over the carbide grains. As the softer matrix is cut away and removed, the load is significantly transferred to hard particles. Incoherent particle wear occurs by two slower processes: abrasion of the hard particles and loss of hard particles by debonding between the matrix and the particles, because of fatigue. For alloyed white cast irons, research indicates that approximately 30 vol% carbides provides the best abrasion resistance.

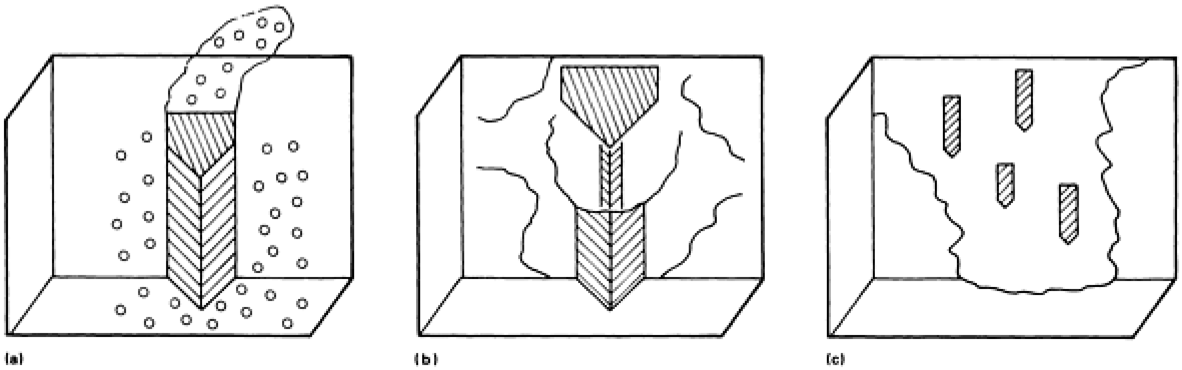


Fig. 6 Effect on abrasive wear when second phase size is varied. (a) Small second phase, easily removed. (b) Large second phase, protection of matrix. (c) Very large second phase, small abrasive channeled to matrix

The relationship between particle size and abrasive grain size is important. Larger abrasive grains tend to create larger wear chips. When incoherent particles are small, relative to the abrasive grains and wear chips, they can be cut out with the matrix, adding little to the abrasion resistance of the material. If the abrasive grains are very small, relative to the hard particles, and the gaps between particles are large, then the grains are able to undermine the hard particles, allowing them to fall out or be dislodged by the occasional large abrasive grain.

The particle characteristics that work best for wear protection are hard, tough, and blocky. A high hardness value makes them harder to cut. Toughness makes them resistant to breakage. Blocky particles, versus those that are plate- or rod-shaped, also reduce crack propagation and breakage.

With the advent of new advanced materials, the use of reinforced composites is becoming more common. These materials are subjected to abrasive wear in many applications. Factors that affect the abrasive wear of these materials include the orientation, size, modulus of elasticity, relative hardness, and brittleness of the second phase. The simplified model of Zum Gahr (Ref 14) (Fig. 7) shows the effects of varying these parameters.

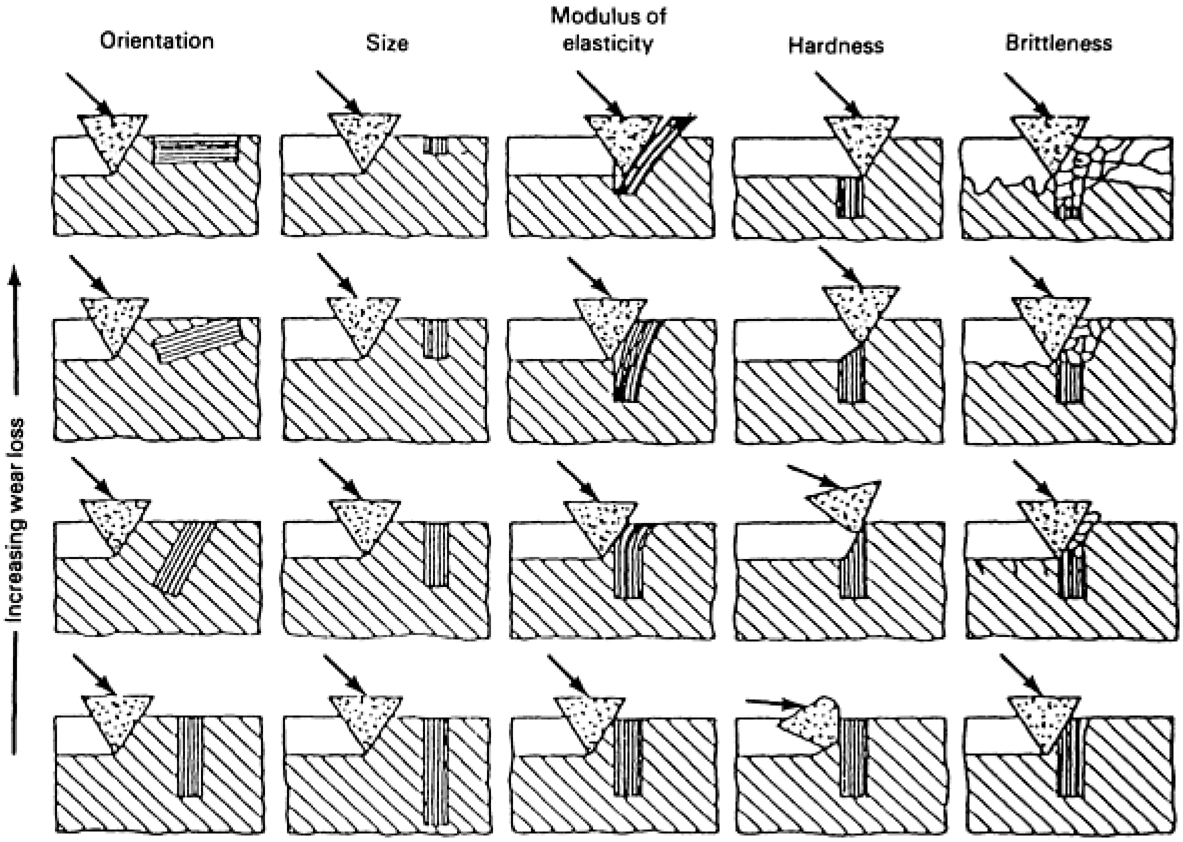


Fig. 7 Effect of orientation, size, elastic modulus, hardness, and brittleness of second phase on abrasive wear. Source: Ref 14

It has been found that a reinforcing second phase lying parallel to the surface is more easily removed than one that is anchored perpendicular to the surface. Also, when the size of the second phase is small relative to the abrasive groove depth, the second phase has little or no beneficial effect. Because most reinforcing additions have a high modulus of elasticity, a matrix with a low modulus will tend to debond at the interfaces and lead to pull-out and abrasive loss. In some metals, such as alloyed white cast irons, if the second phase is harder than the matrix, then the hard phase will protect the matrix. Lastly, brittle materials tend to crack and chip to a larger area than the cross section of the abrasive grain doing the damage. An impressive amount of current research on wear-resistant materials is focusing on advanced composites.

Effect of Environment on Abrasive Wear

In addition to the properties of a material, the environment affects wear. As stated earlier, abrasion loss rates are not intrinsic to a material. Environmental factors that effect abrasive loss include, but are not limited to: the type of abrasive and its characteristics, temperature, speed of contact, unit load of the abrasive on the material, humidity, and corrosive effects, each of which is discussed below.

Abrasive. In the simple model of abrasive wear previously developed, differences in abrasives have been included in the constant and mostly ignored. However, changing the abrasive will change the wear rate. The effect of critical angle has already been discussed, but other abrasive characteristics will also contribute. Among these are hardness, toughness, and size of the abrasive.

The hardness of the abrasive particles is important to the rate of abrasion of the subject material. As the hardness of the abrasive exceeds that of the wear material, abrasive wear typically becomes much worse (Ref 15, 16), as shown in Fig. 8. As the abrasive hardness exceeds the hardness of the material, it is able to penetrate the surface and cut/remove material without having its cutting edges broken or rounded.

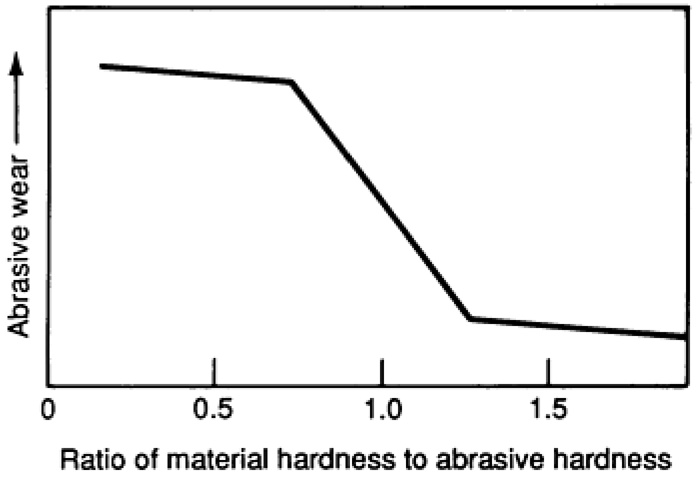


Fig. 8 Effect of abrasive hardness, relative to material hardness, on abrasive wear

Figure 9 shows the hardness of typical minerals and alloy microconstituents. According to the above theory, hematite ore would cut pearlite, but not martensite. This also explains the great advantage of chromium white irons, which have a hardness greater than most common minerals.

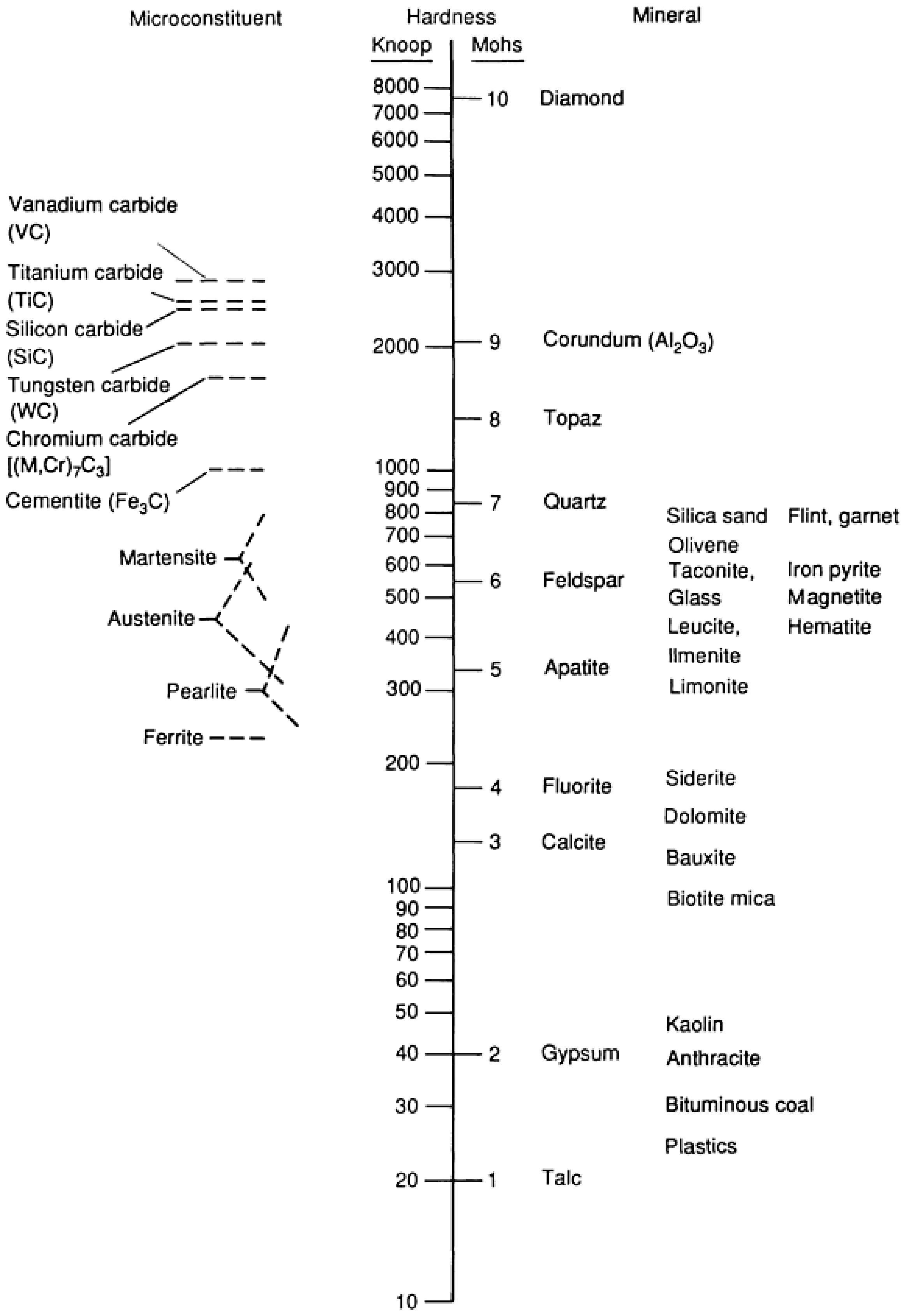


Fig. 9 Hardness of some minerals and alloy microconstituents

The shape of the abrasive particle is important, because it influences the shape of the groove produced in the material. It also influences the contact load and the transition from elastic to plastic contact. Experiments have confirmed that less wear occurs when materials are abraded by rounded, rather than sharp, particles.

The toughness of the abrasive particles is an important factor during abrasion. Material loss will increase when the toughness of the abrasive increases. Avery (Ref 17) gives examples of wear on several white irons and a steel when subjected to two different abrasives of the same hardness (Fig. 10). Although chert and silica both have the same hardness (Mohs 7), the chert, which is the tougher mineral, caused two to three times the wear generated by the silica.

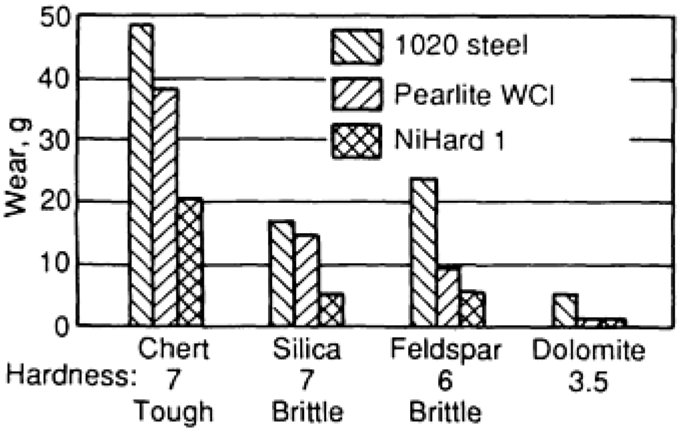


Fig. 10 Effect of toughness of minerals on wear, using ABEX wet abrasion test. Source: Ref 17

Temperature. It might be expected that abrasive wear would increase as the temperature rises, because the hardness and yield strength decrease. Instead, for aluminum and copper (Ref 18), when the temperature was increased from ambient to 673 K, very little change in the abrasive wear rate was observed. It has been proposed that the reason for this small change is that during abrasion, small areas are adiabatically heated. At higher initial temperatures, the metal flow stress is reduced. This results in less heating in the material during the abrasion process. The end result is that areas around the material that is being removed have a similar temperature, independent of starting temperature, and similar abrasion rates.

Speed of Contact. The rate of abrasive wear has been found to slightly increase with increasing speed in the range from 0 to 2.5 m/s (0 to 8.2 ft/s). This increase in wear may be attributable to frictional heating. The effect is small, because all of the abrasion occurs in a near-adiabatic process. This should result in nearly the same peak temperature rise, independent of speed, for the tiny volume of material where the asperities are removing the material.

Load. Abrasive wear has been shown to be proportional to load, following the Archard equation. However, this proportional effect breaks down when the load is high enough to fracture the abrasive particles. If the forces do fracture the abrasive particles and create new sharp points, wear can increase. If the abrasive particle points are rounded, wear will decrease.

Humidity. The effect of atmospheric humidity on abrasive wear is far from clear, and contrary results exist. Larsen-Basse (Ref 19, 20) studied the effect of atmospheric humidity on abrasive wear for a variety of pure metals and steels. When using SiC abrasive, wear usually increased with increasing humidity, up to 65% relative humidity. This increase is attributed to a moisture-assisted fracture of the SiC abrasive particle, which resulted in fresh sharp edges to cut into the surface of the material.

An additional cause for increased wear was the Rehbinder effect (Ref 21). The Rehbinder effect is a chemo-mechanical mechanism in which chemical impurities, such as ions or atomic hydrogen, modify the fracture properties in a crack root, leading to increased wear rates.

Mercer (Ref 22) found a much different result, where, with increasing humidity, abrasive wear decreased for iron and mild steel, stayed constant for titanium, and increased for copper. With contrary results such as these, the effect of humidity is still not clear.

Corrosive Effects. Abrasive wear is often enhanced by corrosive conditions, particularly a low pH. A synergism often occurs between abrasive wear and corrosion. The abrasion creates fresh surfaces that rapidly corrode, and the normally protective corrosion layer is removed by abrasion. Using a laboratory abrasives slurry apparatus. Madsen (Ref 23) demonstrated that the synergism of an abrasive and a corrosive component could be twice that of individual components added together. In a grinding study, Tylczak (Ref 24) showed that grinding in acid waste water increased the wear rate by about twice that of grinding in tap water.

Theory

A number of equations have been used for correlations between wear and other properties. The Archard equation for a relationship of wear with hardness has already been introduced. Khrushchov (Ref 25) demonstrated the correlation with hardness and has proposed an empirical correlation with elastic modulus of the form:

$$W = \frac{k_5}{E^{1.3}} \quad \textbf{(Eq 6)}$$

where E is the elastic modulus.

For pure metals, relationships were found between wear and energy of melting, the combination of atomic weight and Debye temperature, and the combinations of melting points divided by atomic volume. All of these relationships measure interatomic cohesion. However, a fundamental understanding of abrasive wear has not yet been developed from fundamental theories.

Most current theory is based on the concept that abrasion is the process of scratching. Furthermore, most theories simplify the tip of the scratcher as a sharp come. The theories then go on the explain the effect of said come sliding across the surface of a specimen. The Archard equation, with small modifications, is still widely used as a starting point for the development of more complex equations. The more successfully the models deal with "real" complications, the more useful they will be.

One example of an extended Archard mode proposed by Zum Gahr (Ref 26) has a factor that accounts for the proportion of displaced material to removed material. He defines a factor f_{ab} as:

$$f_{ab} = \frac{A_v - (A_1 + A_2)}{A_v} \quad \textbf{(Eq 7)}$$

where A_v is the cross-sectional area of the wear groove, and A_1 and A_2, combined, are the cross-sectional area of the material displaced to the sides of the groove. For perfect cutting, this term is 1, and for pure plowing, with no material removal, the term is 0. The equation for wear, in this case, is:

$$W = f_{ab} A_v d \quad \textbf{(Eq 8)}$$

Factors that will reduce f_{ab} give greater wear resistance without requiring a modification of material hardness. In other words, a material with a greater ability to deform will plow, rather than cut. This follows the results from Khrushchov, where the pure materials, which have a large capacity for deformation, had greater abrasion resistance than alloy steels of a given hardness.

Materials

Material selection for abrasion resistance is not an exact science. Although there seems to be a counter example for every example, there are families of materials that have demonstrated good resistance. They are typically hard materials that resist scratching and include ceramics, carbide materials, alloyed white cast irons, and hardened alloy steels. In addition, these and other materials can be put, or formed, on the surface of many less-abrasion-resistant materials by welding, plasma spraying, flame spraying, electroplating, diffusion, and other techniques.

Ceramics. Many ceramics show outstanding abrasion resistance, because of their high hardness relative to the abrasive particle. One of their weaknesses, however, is an inability to withstand impact, that is, they have a low fracture toughness. At high unit loads, ceramics are also subject to microfracture, which can lead to high wear loss rates. Alumina, zirconia, and silicon nitride, among others, are currently used for wear-resistant applications. Ceramics use will increase as their limitations are further understood and their prices decrease.

Metals, because of their ability to withstand general abuse, will continue to be widely used in abrasive wear applications. As a general guide to metals that will resist abrasive wear, it has been found that:

- High hardness is the primary requirement
- Abrasion resistance tends to increase with additions of carbide-forming metals.
- Carbides are useful additions to metals when they are large in relation to the size of the abrasive.

Alloyed white cast irons are the most abrasion-resistant iron-base alloys. Their resistance to abrasion is due to the formation of carbides during solidification. Unfortunately, they have limited impact resistance compared to steels, are very difficult to machine, and are not weldable, which limits their industrial applications. The families of white irons include high-chromium, chrome-moly, nickel-chrome, and pearlitic white irons. The compositions of typical white cast irons are given at Table 1.

Table 1 Alloy composition of abrasion-resistant white cast irons

Common name	Class	Type	ASTM designation	Composition, wt%						
				C	**Mn**	**Si**	**Ni**	**Cr**	**Mo**	**Cu**
NiHard I	I	A	Ni-Cr-Hc	2.8-3.6	2.0 max	0.8 max	3.3-5.0	1.4-4.0	1.0 max	. . .
NiHard I	I	B	Ni-Cr-Lc	2.4-3.0	2.0 max	0.8 max	3.3-5.0	1.4-4.0	1.0 max	. . .
	I	C	Ni-Cr-GB	2.5-3.7	2.0 max	0.8 max	4.0 max	1.0-2.5	1.0 max	. . .
NiHard IV	I	D	Ni-HiCr	2.5-3.6	2.0 max	2.0 max	4.5-7.0	7.0-11	1.5 max	. . .
	II	A	12% Cr	2.0-3.3	2.0 max	1.5 max	2.5 max	11.0-14.0	3.0 max	1.2 max
Cr-Mo	II	B	15% Cr-Mo	2.0-3.3	2.0 max	1.5 max	2.5 max	14.0-18.0	3.0 max	1.2 max
Cr-Mo	II	D	20% Cr-Mo	2.0-3.3	2.0 max	1.0-2.2	2.5 max	18.0-23.0	3.0 max	1.2 max

The higher-chromium alloys are currently favored over the Ni-Cr or pearlitic white irons, because of their improved resistance to impact in the heat-treated condition. The 25Cr alloy is particularly suited to conditions requiring either extra oxidation or corrosion resistance, whereas the 15Cr3Mo alloys should be used where there is need for thick cross sections, because of their higher hardenability.

Steels. For low-alloy steels, abrasion resistance is basically a function of hardness and carbon content. Other additions can be made to allow for full hardening and other mechanical properties. Tool steels of the M and T (high-speed steel), D (high Cr), and part of the A (air hardening) series often form carbides in the structure for additional abrasion resistance. The American Society of Lubrication Engineers (Ref 27) has developed a chart (Fig. 11) for selecting a replacement tool steel when greater wear resistance, or toughness, is required. Austenitic Mn steels are among the toughest of the available materials. The 12Mn type is known as Hadfield steel. These materials are advantageous where the stresses of use can fully work harden surfaces to help resist further loss.

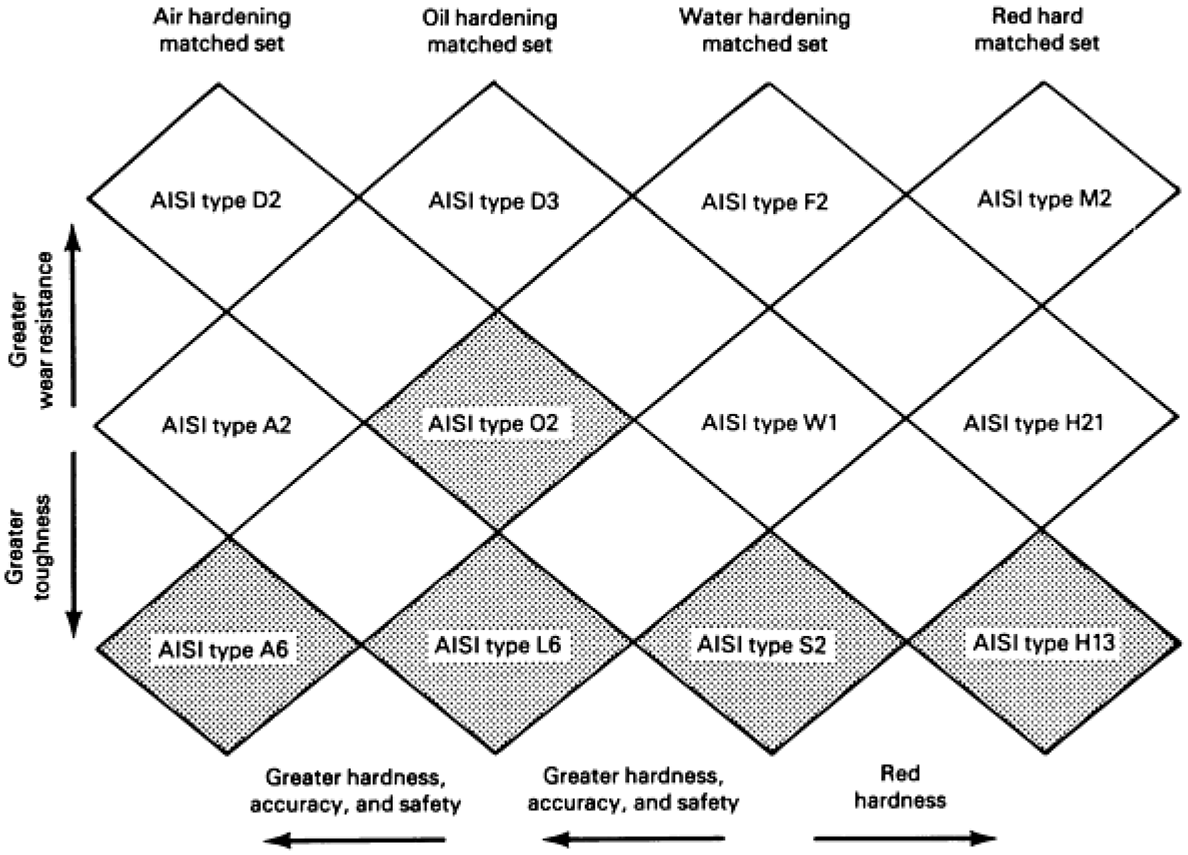


Fig. 11 Basic diagram for matched set method of selecting tool steels. Source: Ref 27

Plastics. It has been observed that plastics have lower wear resistance as the ratio of hardness to elastic modulus decreases (Ref 28). Other work has shown that abrasion resistance increases in both thermoplastics and thermosets, for higher molecular weights of the plastics. These materials are being used for slurry handling because they are able to withstand the abrasion of the small particles of the slurry and are resistant to chemical attack.

References

1. Standard Terminology Relating to Wear and Erosion, *Annual Book of Standards,* Vol 03.02, ASTM, 1987, p 243-250
2. R. Blickensderfer, B.W. Madsen, and J.H. Tylczak, Comparison of Several Types of Abrasive Wear Tests, *Proceedings of the International Conference on Wear of Materials,* American Society of Mechanical Engineers, 1985, p 313
3. A.J. Sedriks and T.O. Mulhearn, The Effect of Work-Hardening on the Mechanics of Cutting in Simulated Abrasive Processes, *Wear,* Vol 7, 1964, p 451
4. T. Sasaki and K. Okamura, The Cutting Mechanism of Abrasive Grain, *Bull. Jpn. Soc. Mech. Eng.,* Vol 12, 1960, p 547
5. K. Kato, Wear Mode Transitions, *Scripta Metall.,* Vol 24, 1990, p 815-820
6. J.F. Archard, Contact and Rubbing of Flat Surfaces, *J. Appl. Phys.,* Vol 24, 1953, p 981
7. G. Sundararajan, A New Model for Two-Body Abrasive Wear Based on the Localization of Plastic Deformation, *Wear,* Vol 117, 1987, p 1-35
8. M.M. Khrushchov and M.A. Babichev, An Investigation of the Wear of Metals and Alloys by Rubbing on an Adhesive Surface, *Friction and Wear in Machinery,* Vol 12, 1958, p 1-13
9. R.C.D. Richardson, The Maximum Hardness of Strained Surfaces and the Abrasive Wear of Metals and

Alloys, *Wear,* Vol 10, 1967, p 353-382

10. P.J. Alison and H. Wilman, The Different Behavior of Hexagonal and Cubic Materials in their Friction, Wear and Work Hardening During Abrasion, *Brit. J. Appl. Phys.,* Vol 15, 1964, p 281-289
11. R.P. Steijn, Friction and Wear of Single Crystals, *Wear,* Vol 7, 1964, p 48-66
12. T.E. Fischer, M.P. Anderson, and S. Jahanmir, Influence of Fracture Toughness on the Wear Resistance of Yttria-doped Zirconium Oxide, *J. Am. Ceram. Soc.,* Vol 72 (No. 2), 1989, p 252-257
13. J.H. Tylczak and R. Blickensderfer, Influence of Composition on Abrasive Wear for Several Binary Systems, *Proceedings of the International Conference on Wear of Materials,* American Society of Mechanical Engineers, 1987, p 645-651
14. K.-H. Zum Gahr, *Microstructure and Wear of Materials,* Elsevier, 1987
15. H. Wahl, Verschleissprobleme im Braunkohlenbergbad, *Braunkohle Wärme Energ.,* Vol 5/6, 1951, p 75-87
16. G.K. Nathan and W.J.D. Jones, Influence of the Hardness of Abrasives on the Abrasive Wear of Metals, *Proceedings of the Institution of Mechanical Engineers,* Vol 181 (No. 30), 1966-67, p 215-221
17. H.S. Avery, "The Measurement of Wear Resistance," Case Report 340-10, Dept. Report 9-AE-134, American Brake Shoe Company, 1961
18. S. Soemantri, A.C. McGee, and I. Finnie, Some Aspects of Abrasive Wear at Elevated Temperatures, *Proceedings of the International Conference on Wear of Materials,* American Society of Mechanical Engineers, 1985, p 338
19. J. Larsen-Basse, Influence of Atmospheric Humidity on Abrasive Wear--I. 3-Body Abrasion, *Wear,* Vol 31, 1975, p 373-379
20. J. Larsen-Basse, Influence of Atmospheric Humidity on Abrasive Wear--II. 2-Body Abrasion, *Wear,* Vol 32, 1975, p 9-14
21. M. Ciftan and E. Saibel, Rehbinder Effect and Wear, *Proceedings of the International Conference on Wear of Materials,* American Society of Mechanical Engineers, 1979, p 659-664
22. A.D. Mercer and I.M. Hutchings, The Influence of Atmospheric Humidity on the Abrasive Wear of Metals, *Proceedings of the International Conference on Wear of Materials,* American Society of Mechanical Engineers, 1985, p 332-337
23. B.W. Madsen, Measurements of Erosion-Corrosion Synergism with a Slurry Wear Test Apparatus, *Proceedings of the International Conference on Wear of Materials,* American Society of Mechanical Engineers, 1987, p 777-786
24. J.H. Tylczak, D.J. Singleton, and R. Blickensderfer, Wear of 12 Alloys During Laboratory Milling of Phosphate Rock in Phosphoric Acid Waste Water, *Miner. Metall. Process.,* Aug 1986, p 187-190
25. M.M. Khrushchov, Principles of Abrasive Wear, *Wear,* Vol 28, 1974, p 69-88
26. K.-H. Zum Gahr, Modeling of Two-body Abrasive Wear, *Wear,* Vol 124, 1988, p 87-103
27. H.A. Garner, Dry Sand/Rubber Wheel Abrasion Wear Testing of the Carpenter Matched Tool and Die Steels, *J. Am. Soc. Lubr. Eng.,* June 1982, p 359-364
28. V.A. Bely, A.I. Sviridenok, M.I. Petrokovets, and V.G. Savkin, *Friction and Wear in Polymer-Based Materials,* Pergamon Press, 1982

Polishing Wear

Leonard E. Samuels, Samuels Consultants

Introduction

THE TERM "POLISHING WEAR" is used to describe interactions between two solids that remove material from, while at the same time producing a polished finish on, the surface of at least one of the two. This definition is not, however, a precise one because the surface condition known as a "polish" cannot be defined quantitatively. The term is, in fact, merely one of common usage describing a surface that reflects light brightly and that produces a clear image of distant objects in the manner expected of a mirror.

The light reflected from real surfaces, which inevitably are rough on a microscopic scale, consists of two components, one known as "specular" or "regular" reflection and the other as "diffuse" reflection. The specular component consists of rays reflected over a narrow range of angles, the same angle as the angle of incidence; these rays are required for the production of sharp reflected images. The diffuse component consists of rays reflected or scattered over a range of angles centered about the angle of specular reflection. Light does not just bounce off a reflecting surface. The incident beam is diffracted from many centers on the surface, and the emergent beam is the net result of the interactions between these diffracted rays. Specular reflection results when the diffracted rays reinforce one another at the appropriate angle and cancel one another at all other angles. For regular parallel grooves in an ideal flat surface, this phenomenon is possible only when the spacing of the centers at which diffraction occurs is less than the wavelength of the incident light (about 500 nm for visible light). The analysis of the reflectivity of actual engineering surfaces, such as abrasive-machined surfaces, however, is complicated by their complex topography, with the spacing, depth, and contour of the roughness grooves varying over a wide range. Theoretical and experimental studies (Ref 1) indicate that the variation in the intensity of specular reflection can best be related to a probability distribution of the heights of these roughness grooves. On the other hand, the spread of the diffusely reflected light depends on the distribution of the slopes of the roughnesses.

A surface has to specularly reflect a large proportion of the incident light before it would be regarded as being "polished," but the degree of reflection and the angle of viewing are very subjective requirements. Thus, there can be no sharp distinction between polished and nonpolished surfaces, as defined by either their reflectivity or their topography. In general, the roughnesses in a reasonably well-polished surface prepared with abrasives consist of grooves spaced ~100 nm apart having a side slope of ≥100° and a length at least several orders of magnitude greater than the width of the grooves.

Because polished surfaces and unpolished surfaces differ only in degree, it may be concluded that polishing wear should only be considered as a special case of certain basic wear processes. This is true only to a degree, and thus this article will concentrate on the circumstances that can generate the required type of finish. Nevertheless, unique wear mechanisms are involved and will have to be considered. Unfortunately, polishing wear encountered in complex engineering systems has not been extensively studied; however, simple systems operated with the deliberate intent of producing a specific polish have been studied, but only to a limited extent. Consequently, the present discussion will necessarily have to be based on these limited studies of deliberate polishing processes, making adjustments as needed to recognize that the objective of these studies of deliberate polishing and of polishing wear have opposing goals; the objective of one is to maximize material removal rate (MRR), and the objective of the other is to minimize the material removal rate. Only materials with high intrinsic reflectivity, the majority of which are metals, need to be considered.

Polishing Wear with Abrasives

Abrasive action is one of the most common causes of polishing wear. It is one of the few types of wear processes that can easily occur on a fine enough scale to generate surfaces with the required topography. Polishing wear under these circumstances can be regarded as a special case of abrasive wear (see the article "Abrasive Wear" in this Section), which typically produces diffuse reflecting surfaces. Consequently, what needs to be discussed here are the conditions under which an abrasive wear system might be expected to generate a surface with a fine topography.

Property Requirements of an Abrasive

Typically, an abrasive is a hard material in the form of small particles (often called grits) bounded by irregularly shaped surfaces that meet at sharp edges and points (Fig. 1). For commercial abrasives, the shape of the grits does not differ greatly, and it is the shape of their points and edges that is of primary importance; the diameter of the particles themselves is of secondary importance. This is demonstrated by the following example. An abrasive with a nominal particle diameter of, say, 10 μm (400 μin.) can be used to produce a good quality polish on many metals. The grooves produced on the workpiece surface with such an abrasive would certainly be <0.5 μm (<20 μin.) wide. Contacting grits can penetrate to only a small fraction of their diameter when producing such groves (Fig. 1), and the penetrating portion must comprise a point or, less likely, an edge. Thus, only the points and edges of abrasive grits actively contribute to a polishing process, and it it the shape of these points or edges that is the determining factor. The primary function of the rest of the individual grit is to support the contacting regions in the same way that the shank of the lathe tool supports the cutting tool. Nevertheless, the dimensions of the particles can have a secondary influence if the maximum number of particles that can be packed together in a given area and, less obviously, the means by which the abrasive particles are held against the workpiece surface can be determined.

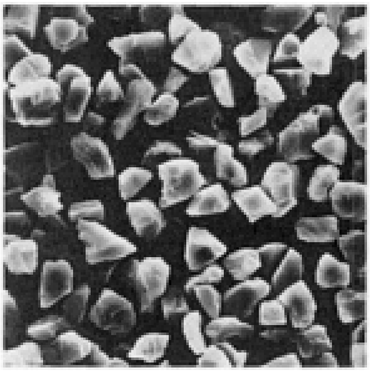


Fig. 1 SEM micrograph of monocrystal diamond polishing abrasive that has a mean particle size of 5 μm (200 μin.). Grit shape is typical of that of most types and grades of abrasives on the market. Width of groove or scratch generated by abrasive is $\approx$0.1 times the grit diameter. 750×

Abrasives are typically considered to be hard materials (for example, silicon carbide, aluminum oxide, and diamond). Nevertheless, the hardnesses of materials conventionally classified as abrasives vary over a considerable range (Table 1), from a value comparable to that of the harder metals (for example, silica, the abrasive most commonly present in natural wear systems) to that of diamond, the hardest material known. In fact, the only requirement of an abrasive from this point of view is that its yield strength be great enough relative to the workpiece to indent the workpiece without the abrasive deforming significantly itself. Strength (that is, hardness) beyond this level of performance is of no benefit.

Table 1 Typical values of hardness and toughness of selected abrasives and metals

Material	Hardness, HV	Fracture toughness[(a)]	
		MPa $\sqrt{m}$	**ksi $\sqrt{in.}$**
Abrasives			
Silica (SiO_2)	850	1.1	1.0
Zirconia (ZrO)	1150	[(b)]	[(b)]
Emery (Al_2O_3·Fe_3O_4)	1400	[(b)]	[(b)]
Alumina (Al_2O_3)	2000	2.2	2.0
Silicon carbide (SiC)	2100	2.5	2.3
Cubic boron nitride (CBN)	2750	5.7	5.2
Diamond (C)	8000	[(b)]	[(b)]
Metals			
Steel			
1.2% C, hardened	950	[(b)]	[(b)]
0.4%			
Hardened and tempered	400	70	64

Normalized	190	50	46
Ferrite phase	75	(b)	(b)
Cementite phase	1100	(b)	(b)
Aluminium alloy, age hardened	135	30	27

(a) Source: Ref 2.

(b) No data available

Indentation hardness values, or ordinarily determined, the valid indicators of the relative properties of the abrasive and the workpiece in this respect, and a sample theoretical analysis developed by Tabor (Ref 3) has indicated that the hardness of the abrasive should be two to three times that of the workpiece if the abrasive is not to deform noticeably during indentation. However, an abrasive does not have to be quite that hard to indent the workpiece to some extent. Even if the abrasive itself is deformed a little in the process, it is still capable of removing material with at least a degree of effectiveness that would generate workpiece wear. In fact, one experimental investigation (Ref 4) has suggested that an abrasive need only be slightly harder than the workpiece itself to effect some abrasive wear. The material-removal efficiency was then found to increase rapidly with the increase in the hardness ratio up to a value of 1.5 but leveled off at hardness ratios >1.5. Unfortunately, neither the mechanisms by which the material was being removed nor the effect on surface finish was determined in this investigation. Therefore, these results cannot be precisely interpreted. Nevertheless, the general point has been established that a so-called abrasive does not necessarily require a high hardness top effect wear but only that the hardness of the abrasive must exceed the hardness of workpiece material. Thus, particles of a hard metal could produce abrasive wear in many comparatively soft metals. Protuberances on the surface of a hard metal could also generate abrasive wear. However, hard materials such as the classical abrasives are required to produce wear in the harder alloys used in engineering practice.

A difficulty arises when applying to polishing wear those concepts that were developed with larger-scale processes such as abrasive wear and abrasive machining (grinding). Indentation hardness tests, even the microindentation hardness tests on which the concepts of abrasive wear have been based, are carried out on a much larger scale than the processes occurring during polishing wear. The question arises as to whether these abrasive processes are adequately representative under these circumstances. It was originally believed that the principle of similarity, on which hardness comparisons rely, no longer applied to indentations smaller than a certain size. However, the best evidence currently available indicates that the principle does indeed apply down to the smallest conceivable indentations (Ref 5, 6, 7). The comparisons are still actually valid only if the workpiece material is fairly homogeneous over the entire range of indentation sizes specified. However, many metallic alloys are not homogeneous at the microscopic level at which polishing wear is produced. These metallic alloys can contain phases whose mechanical properties are considerably different from the matrix phase and whose size can vary over a wide range depending on the composition, mechanical properties, and thermal history of the alloy. It is quite possible for such a phase to be small enough relative to the scale of the grinding or abrasion process that it will not have a significant effect but be large enough on the scale of polishing wear to have a major effect on the performance of some types of abrasive grits. In this case, to act as an abrasive, a material would have to be matched to the characteristics of the second phase for polishing wear but not necessarily so for abrasive wear.

It has been recognized in other contexts (Ref 8) that it is convenient to divide abrasion-sensitive materials into two categories:

- All single-phase metals and alloys and those multiphase alloys in which either the properties of the additional phases do not differ significantly from the matrix phase or the particles of the additional phases are small relative to the grooves made by the abrasive particles.
- Multiphase alloys in which one or more of the minor phases is much harder than the matrix phase and is large relative to the grooves made by the abrasive grits

As we have just seen, it is also convenient to categorize materials in the same way as for polishing wear. The difference is the presence of much smaller particles of a hard phase may require that an alloy be included in the second category.

Fracture Toughness. An abrasive may easily meet the hardness ratio criterion mentioned earlier and still not be applicable for a wear system because it fractures easily, either by crushing in compression or by fracturing in tension (due to the bending forces imposed on a contacting grit point as it plows across the workpiece surface). Fracture in bending is

the more likely cause of point fracture because abrasives have poor fracture toughness. That is, abrasives are prone to catastrophic failure when stressed in tension in the presence of stress concentrators. Stress concentrators in the form of notches and cracks inevitably are present in abrasive grits. Fracture toughness can be quantified even in brittle materials. The fracture toughness values for all abrasives are low, but the values differ in a manner that is not necessarily related to hardness. Silica, for example, is the least tough of the common abrasive materials even though it is the softest of them all; diamond, on the other hand, is the hardest of all abrasives and yet probably is the toughest. All common abrasives are significantly less tough than general engineering materials are judged by the values in Table 1, implying that common abrasives would fracture before the workpiece when the two are stressed in contact. The abrasive would then gradually be removed from the system, more so with some abrasives than others. The solution is less clear-cut, however, when brittle phases are present in the workpiece material because the brittleness of some of these phases might approach that of the abrasives. A moderating factor, however, is that small particles of brittle phases are more ductile than expected when supported by a surrounding ductile matrix, but minimal detailed information is available on this phenomenon.

Nevertheless, these comparisons of fracture toughness predict a deterioration in the effectiveness of abrasives in wear systems. The comparison can be expected to be reasonably valid on a large-scale system typical of grinding machining, for example, but it may not be valid on the small-scale system in which polishing wear occurs. This is because small volumes of material are being deformed by an unusual and complex compressive stress system. The work of Bridgman (Ref 9) showed that materials that are brittle in bulk can react in a ductile manner under these circumstances. This certainly appears to be the case for the workpiece material but whether it is also the case for the contacting points of abrasive grits is a matter for conjecture. Under such conditions, it would mean that certain abrasive materials might have a more serious effect on polishing wear than would be expected from their reaction in larger-scale processes.

Melting Point. It has been suggested in the past that the melting point is one physical property that has to be taken into consideration when assessing whether a particular material is likely to act as a polishing abrasive. It was believed that a material had the ability to polish only if its melting point was higher than that of the workpiece material (Ref 10). This conclusion was based on polishing trials with a reasonably wide range of abrasive-workpiece combinations, each of which initially supported this conclusion. However, trials subsequently carried out with a wider range of combinations (Ref 11) unearthed many systems that did not conform to the initial results (Fig. 2). In the combinations investigated in the original experiments of Bowden and Hughes (Ref 10), black dots were found only above the line of equal melting points (Fig. 2). Subsequent experiments by Rabinowicz (Ref 11) produced the combinations that are found below the line. Therefore, the concept of the significance of relative melting points can no longer be regarded as being valid. On the other hand, it is true that hard materials tend to have high melting points. Thus, a general if indirect correlation can be made between polishing ability and the melting point of an abrasive, although no reliance can be placed on the correlation in specific instances.

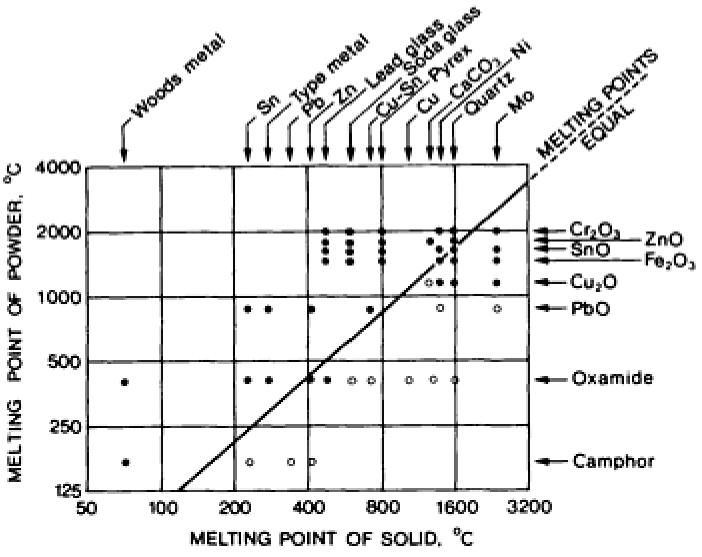


Fig. 2 Correlation between polishing ability and the relative melting points of the polishing powder and the workpiece solid. Solid dots indicate abrasive powder and solid combinations for which polishing does occur, whereas open dots indicate powder-solid combinations that do not generate a polished surface. Source: Ref 10,

Parameters Required to Generate Specularly Reflecting Topographies

Experience indicates that the included angle of potential contacting points of abrasive grits is typically $\geq$100°, and indentations by points as obtuse as this can be taken to be directly analogous to those made by the Vickers-type or Knoop-type indenter, which has an included pyramidal angle of >100° (Ref 12). Consequently the applied force required to produce a static indentation of a particular grit size can be calculated if the true surface hardness of the workpiece is known and the abrasive: workpiece hardness ratio criterion (see the section "Property Requirements of an Abrasive" in this article) has been complied with. The groove produced when an indented grit is then moved across the workpiece surface by a lateral force is known to be narrower than the static indentation but only by ~10% (Ref 13), a difference that can be ignored for the purposes of this discussion. On this basis, an approximate estimate can be made of the largest force that can be applied to contacting grits if they are to produce grooves that are narrow enough to generate a polished appearance (that is, if they are to produce grooves narrower than 50 to 100 nm). A limiting force of ~0.01 N (1 gf) is indicated for a workpiece hardness of 1000 HV and limiting force of 0.002 N (0.2 gf) for a hardness of 100 HV. The next step is to determine under what conditions the applied loads can be limited to this order to magnitude.

In the final analysis, the force applied to an abrasive particle is determined by the elastic constants of the entire system that holds the abrasive grits against the workpiece surface. This system consists of a number of units linked in a closed loop (namely, the grits; the cement, or any other device that holds the grits in place; the mechanical components that connect the workpiece and the abrasive; and the workpiece itself). Grinding machines of the type used in machine shop practice are representative of so-called hard elastic systems (that is, small changes in displacement between the grinding wheel and workpiece cause large changes in the force exerted between the two, principally because the mechanical components of the grinding machine simulate a very stiff spring). Consequently, such a machine can be expected to produce a polished surface if, and only if, the depth of indentation of the contacting grits can be kept appropriately low and if the whole system can be controlled precisely so that the penetration depth is maintained at the low value throughout the machining cycle. This can be done (Fig. 3), but only under special conditions in precision machines (Ref 14). The grinding system required to produce such a highly polished finish must be very rigid and the depth of cut of the grinding passes has to be kept to an unusually small value to ensure that the surface grooves formed are shallow enough and closely spaced enough for specular reflectivity. Even then, a high polish can be obtained only on hard materials. These conditions are not likely to be achieved in an elastically-hard wear system but at least the principle is established that all which is required to produce a "polished" surface is to restrict the width and depth of the grooves generated on the surface. However, it also suggests that wear which yields polished surfaces is not likely to occur in elastically hard systems.

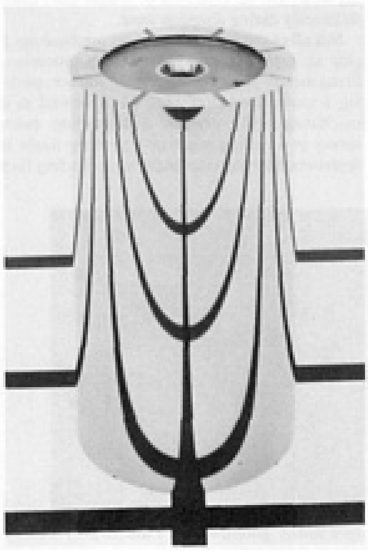


Fig. 3 Hard steel cylinder whose surface was ground to a polish using precision toolroom practices. The cylindrical surface exhibits a high degree of specular reflectivity, as indicated by the clear and sharp reflections of the grid lines ruled on the plane on which the cylinder rests.

Cloth laps charged with a fine abrasive suspended in a liquid, such as those used in metallographic practice, are examples of soft elastic systems. The workpiece surface actually contacts the fibers of the cloth, either individual fibers or groups of fibers incorporated into a thread, and each fiber simulates can elastically soft cantilever spring (Fig. 4). The fibers are typically ~20 μm (~800 μin.) in diameter and made from materials with a low elastic modulus so that they can exert only a very small reactive force when distorted. Moreover, this reactive force would change little with variation in the distortion. Increasing the force applied to a workpiece causes it to sink into the cloth, thus increasing the number of fibers contacted but not markedly increasing the force applied by each component. The local force system is, in effect, self correcting, which permits comparatively easy control of surface finish. The dimensions of the cloth fibers, the method by which they are held or woven together, the elastic modulus of the fiber material, and the number of abrasive grits held in contact by each fiber are consequently the factors that determine surface finish as well as, but usually to a lesser extent than, the force applied to the workpiece. This can be regarded as a more plausible, if perhaps excessively idealized, model of the conditions that are likely to cause polishing wear.

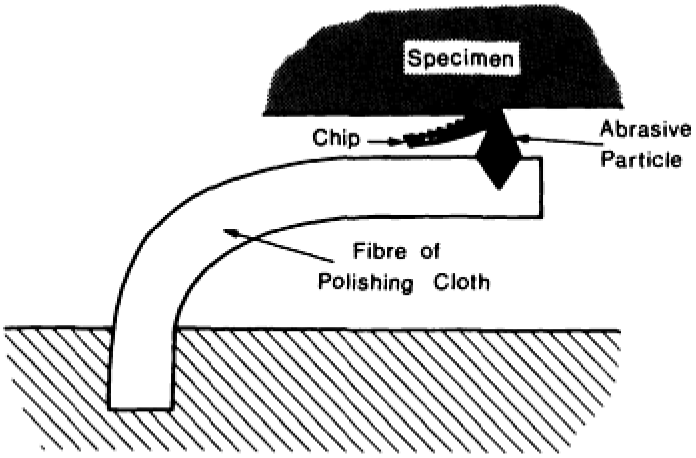


Fig. 4 Schematic showing one method by which an abrasive grit can be held in a fixed position to ensure the application of a very small indenting force. This indenting force is determined by the elastic characteristics of a textile fiber that simulates a very soft cantilever spring. Source: Ref 15, 16

However, extended grooves of the type typically present on specularly reflecting surfaces will be formed only if the grits are held fixed in one orientation in space for a specified duration (Fig. 5). Loose tumbling grits produce angular indentations (Fig. 5b) that are conducive to diffuse the reflectivity. It is difficult to see how grits, initially added as a powder, can be held fixed in this way when the abrasive is present on an elastically soft support such as a cloth, but there are several possibilities (Ref 16, 17):

- Grits can become embedded in the fibers of a cloth (Fig. 6a), provided that their diameter is less than the fiber diameter
- Groups of grits can be attached to the fibers by a viscous liquid or a waxy solid (Fig. 6b); this is likely only when the grits are much smaller than the fiber diameter
- Grits can be snared between tangled fibers (Fig. 6c); this is likely only when the grits are much larger than the fiber diameter

The second arrangement is most likely to result in the highest material removal rate, whereas the third arrangement results in the lowest material removal rate.

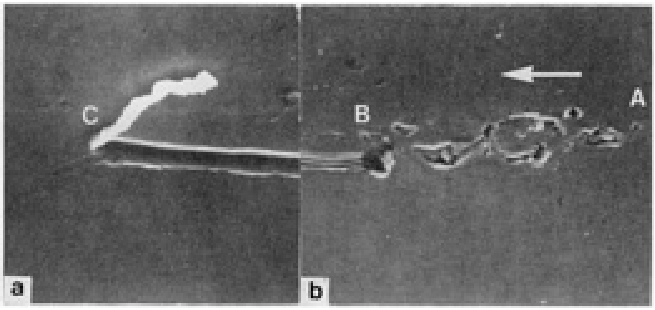


Fig. 5 SEM micrographs of a track on a prepolished copper surface generated by a 1 μm (40 μin.) diamond abrasive grit. (a) 7000×. (b) 5000×. The grit, moving in the direction of the arrow (b) relative to the specimen, first contacted the specimen surface at about point A and made an angular indentation at that location. The grit then tumbled across the specimen surface making additional angular indentations as the numerous edges and corners of the grit successively came into contact. At location B, the grit became fixed in position, perhaps by either one of the methods shown in (a) (b). Thereafter, it formed an extended groove, the end of which is shown in (a). A ribbon of metal that has all the characteristics of a machining chip is attached at the termination of the groove at location C. This groove was clearly machined out of the specimen surface.

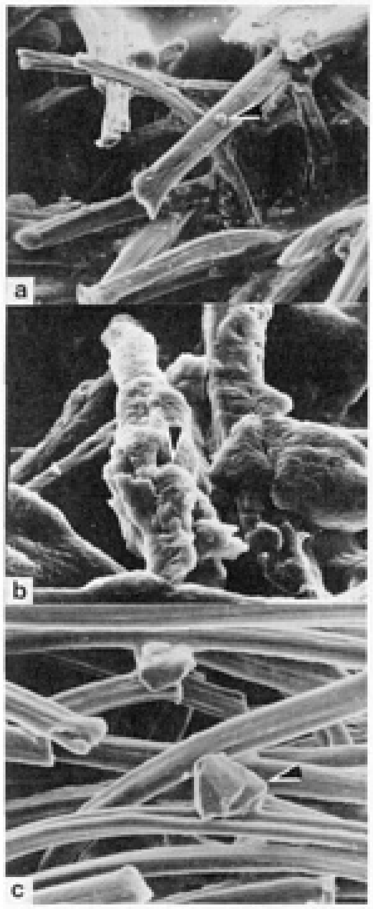


Fig. 6 SEM micrographs showing different methods of holding abrasive grits on a cloth in a fixed position and orientation. (a) Isolated grits (indicated by arrow) embedded in cloth fibers (210×). (b) Numerous grits (indicated by arrow) held to fibers by a pasty material (500×). (c) Isolated grits (indicated by arrow) entangled between the cloth fibers (250×). Grits must be held fixed to form an extended groove in a workpiece surface.

It is also possible that the grits could be circulated by turbulences developed in the liquid in which it is suspended and occasionally whisk across the workpiece surface with sufficient momentum to form an extended groove. In such an event,

polishing wear would become a special case of slurry erosion. There is direct evidence that circulation of the grits does occur by this method (Ref 17), but only inferential evidence that extended grooves can be formed by this phenomenon. Indeed, the indications are that they can most readily be formed on harder materials and less readily on soft materials and phases (Fig. 7). Indentations rather than grooves tend to be formed in the latter case (Fig. 7b).

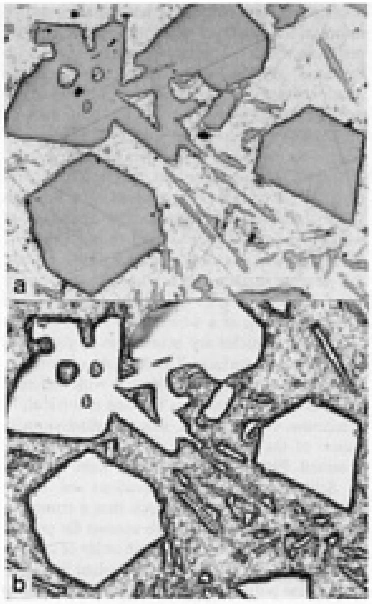


Fig. 7 Optical micrographs of an Al-13Si alloy polished in a dry abrasive and an abrasive entrained in a water slurry. (a) Diamond abrasive applied so that the grits were held fixed to the fibers of a polishing cloth as shown in Fig. 6(b). Alumina abrasive of a grade corresponding to diamond but applied as a water slurry. The fixed diamond grits operated in a cutting mode and cut extended grooves in both the hard silicon phase (angular light gray) and the soft aluminum phase of the matrix. The loose alumina grits, presumably operated in an erosive mode, produced only indentations in the aluminum phase, which consequently has a comparatively poor reflecting matte finish. Both 500×

Mechanisms of Material Removal

The creation of a system of grooves may produce a specularly reflecting surface, but it does not follow that material is removed during the process (that is, that the workpiece actually wears). Little if any information is available on the subject of the material removal mechanisms involved in polishing wear systems. Consequently, it is necessary to make assumptions that are based on the knowledge of other systems, particularly deliberate polishing systems.

Chip Machining Mechanism. The most plausible explanation is that the contacting points of either abrasive-like grits or hard protuberances on another surface simulate tools that produce the grooves by machining out chips. This is the mechanism primarily responsible for the generation of surfaces in ductile materials by abrasive machining processes ranging from industrial grinding to polishing (Ref 15, 16, 18, 19). Evidence of the type illustrated in Fig. 5(a) and Fig. 8(a) indicates conclusively that chip machining is the dominant material removal mechanism in almost all polishing situations. Chip machining is even the dominant mechanism for many materials that are brittle in bulk form (Ref 20, 21). For reasons that are not clearly established and that need not be discussed at this point, these materials are more ductile during indentation when the depth of indentation is less than a critical value, where this critical value can be of the order of magnitude encountered in polishing. Therefore, hard brittle materials or phases may incur wear by the same type of mechanisms as ductile materials during polishing wear, but the materials or phases react quite differently during abrasion wear.

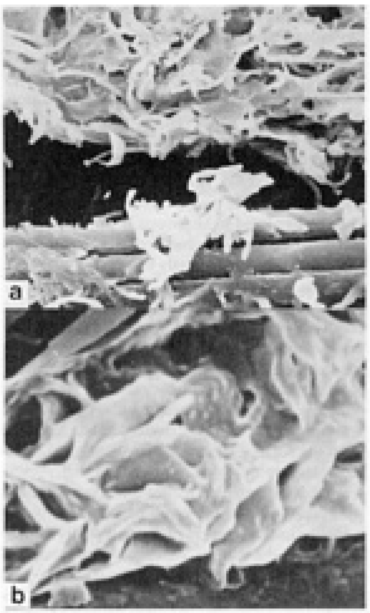


Fig. 8 SEM micrographs of metallic debris remaining on polishing cloths after polishing a Cu-30Zn brass with diamond abrasives. (a) Debris produced during polishing on a comparatively coarse (30 μm, or 1200 μin., mean particle diameter) abrasive. The debris consists almost entirely of threads of metal that have the characteristics of machining chips. 1000×. (b) Platelet-type debris produced during polishing with a very fine abrasive (0.1 μm, or 4 μin., maximum particle diameter). Such platelets are presumed to be generated by a delamination mechanism of material removal and thus are quite different from the debris shown in (a). 5000×. Source: Ref 16

Not all of the contacting points machine out a chip in typical abrasive-machining processes. Some merely plow a groove in the surface, pushing a stem ahead (Ref 18). When viewed as a machining tool, whether a contacting point serves in a cutting mode or a plowing mode is determined by the rake angle of its leading face and the nature of the workpiece material. A point cuts a chip only when its rake angle is more positive than a critical value, a value that is a characteristic principally of the workpiece material (Ref 13, 18). Large numbers of points with a shape suitable for chip cutting are conducive to relatively high rates of material removal, which is desirable for deliberate polishing but clearly undesirable for wear applications. Although it is reasonable to assume that these principles would apply to abrasive polishing, there is no concrete evidence available to confirm this. No information is available on basic parameters (for example, the rake angles of contacting points and the values of critical rake angles) that would be needed for the prediction of wear rates.

The optimum condition from the standpoint of minimizing wear rate is one in which all of the contacting points are operating in a plow mode. Nevertheless, it is still possible that the points would remove material by mechanisms that are of minor importance in abrasive machining but that could be significant in terms of wear. One of these conditions occurs with contacting points that have negative side clearances when serving as machining tools. Contacting points with negative side clearances are commonly encountered in machining practice. Some of the material flowing into the prow established ahead of the contacting point tends to flow towards one side of the groove being produced, thus forming a thin fin at the edge of the groove (Fig. 9). This fin is sure to break off sooner or later. Another potential condition is that the prows themselves sometimes detach by shear fracture at their base (Ref 19). Prows are also likely to detach, partly if not completely, when they reach an external edge of the workpiece. Finally, a prow sometimes may behave as the equivalent of a built-up edge (BUE) on a tool and cause the separation of a small secondary chip when the contacting point itself would not do so (Ref 19). Although each one of these phenomena can be assumed to occur in polishing wear, no information is available about the conditions under which they are likely to occur and, hence, their likely contribution to material removal cannot be quantified.

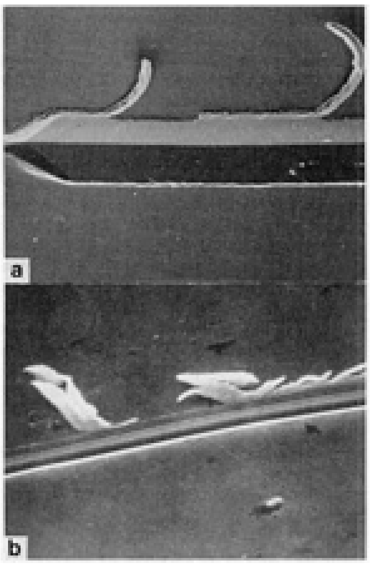


Fig. 9 SEM micrographs of fins formed at the edge of an abrasion groove by a grit having a negative side-clearance angle. (a) End of a groove generated by a model pyramidal point oriented to have a negative side-clearance angle. A portion of the prow formed ahead of the point has flowed into one of the side edges of the groove, developing a fin that periodically separates from the edge. 138×. (b) Abrasion groove formed under actual conditions. An irregular fin has formed at one edge of the groove and is certain to detach, thus constituting a secondary mechanism of material removal. 460×

Delamination Mechanism. An entirely different material removal mechanism from that just described has been observed to occur during certain conditions of deliberate polishing. These conditions involve the use of very fine abrasives (<1 μm, or 40 μin., particle size) and produce surfaces that are brilliantly specular but still composed of systems of parallel grooves. Material removal in this case occurs via the very small labs of material being lifted out of the surface (Fig. 8b), presumably by some form of delamination. Moreover, it seems that this delamination phenomenon is typical, occurring in varying degrees in all abrasive-machining processes from grinding (Ref 22) to polishing (Ref 16); however, this phenomenon is overwhelmed by the chip-cutting mechanisms in the coarser processes. Yet for some as yet unknown reason, the chip-cutting mechanism begins to be phased out at a certain stage of the process as the finish flatness is increased. It eventually ceases, and the delamination mechanism is then left as the sole material removal process. The amount of material removed by the delamination mechanism is likely to be much less than that removed by chip cutting. An adequate explanation for this phenomenon has not yet been developed (Ref 22, 23).

Multiple-Pass Mechanisms. All of the mechanisms discussed to this point assume that the active contacting point is harder than the workpiece (that is, it meets the Tabor criterion); moreover, they assume only a single pass of the contacting point. We have already noted that some commonly occurring abrasive materials cause wear even though they are considered to be comparatively soft by these standards. In addition multiple passes are likely in wear, particularly when material is being removed at a low rate. Kosel and his colleagues (Ref 8, 24, 25) have shown that abrasives can wear harder materials and phases after multiple passes and that material removal can then occur by several mechanisms (for example, delamination and the removal of fragments by fracture). The resultant wear rate is small. The conditions necessary to induce these mechanisms, however, seem to be too severe to be conducive to the generation of specularly reflecting surfaces. Nevertheless, it is expected that multiple passes could enhance the significance of delamination and other secondary processes.

Erosion Mechanisms. It was mentioned in the section "Parameters Required to Generate Specularly Reflecting Topographies" that polishing wear can be caused by an abrasive suspended in a liquid that circulates past the workpiece surface. Normal erosion wear mechanisms can then be expected to operate. The mechanism most likely to occur depends on the momentum of individual grits perhaps being sufficient for the grits to indent into the workpiece surface when they strike the surface and then to sweep across the surface, producing an arc-shaped groove. The grits would either plow or machine out a chip, depending on the shape of their contacting point, typically the same process encountered for fixed abrasives. Small grits and low flow velocities would tend to produce shallow grooves conducive to the generation of a polished finish and comparatively low wear rates. However, it appears that highly reflecting surfaces are likely to be produced by this method only on harder phases; poorly reflecting indented surfaces tend to be produced on soft phases

(Fig. 7b). A further consequence is that adjoining particles or phases of an alloy with differing polishing characteristics are more likely to polish to different levels than when the abrasives operate in a chip-cutting mode. The difference in relief between phases may even be discernible to the naked eye. Moreover, for reasons that are not clear, preferential erosion tends to occur locally at the interface between phases that have markedly different polishing characteristics (Ref 15).

Summary. Assuming that the abrasive material has adequate mechanical properties to withstand undamaged its interaction with the workpiece, it is the shape of the points of the grits and the orientation of the grits as they encounter the workpiece surface that determine their potential to cause wear. The term "shape" refers to the details of the configuration of the small portion of the point that actually contacts the workpiece surface. A perfectly smooth spherical point, for example, might generate a groove in the surface, but it typically would not remove material by machining out a chip because it would present only a highly negative rake angle as a machining tool. On the other hand, a point that contains numerous fracture facets, even if it has a generally spherical form, might be able to do so because some of the fracture facets might have a shape that enables it to act as a machining tool. These facets could be original or they could be the result of fracture in use (Ref 25). However, these cases are contingent on the grit being held in a fixed position in space for a sufficient time to machine a lengthy groove.

Wear Rates

It is a widely accepted belief that polishing wear rates are inversely proportional to workpiece hardness. This view has been implied throughout this article. However, the presumed relationship actually is based on little more than intuition because it has never been supported by quantitative experimental evidence. In addition, theoretical analyses, even the very generalized ones given above, indicate that a simple relationship of this nature is not to be expected.

The width (and hence the volume) of the grooves generated on a polished surface can indeed be expected to be inversely proportional to the local surface hardness of the workpiece. Even this is not necessarily identical to the bulk hardness of the material because it usually would have to be adjusted to account for the effects of surface layer deformation during previous machining and wear and to account for the hardness of individual phases of the microstructure. Even then, a correlation with workpiece hardness would be obtained only if all of the volume of the surface grooves were removed from the surface (that is, if the theoretical limiting wear rate was obtained). An alternative possibility would be that a constant fraction of the grooves was removed for all materials, which is a highly unlikely event. In reality, only a fraction, x, of the groove volume is removed, and this fraction varies considerably $(0 < x < 1)$ with different materials. In fact, the fraction depends on a multitude of factors, many of which are in no way related to workpiece hardness. The fraction cannot be predicted, although experimental determination is sometimes possible. The variation in this fraction is sufficient to negate the possibility of a correlation between wear rate and hardness.

Experimental evidence on the effects of the properties of the workpiece material on polishing rate is scant and confined to metallographic polishing processes. Nevertheless, the evidence that is available confirms the complexity of the matter. Figure 10 explores the possibility of a relationship with workpiece hardness. Figure 10 indicates the existence of an upper boundary for results in which the polishing rate decreases with hardness. Within this boundary, however, the relationship is highly random. A particularly anomalous feature is that the polishing rates of the alloys considered are, with only one exception, higher for their softer base metals.

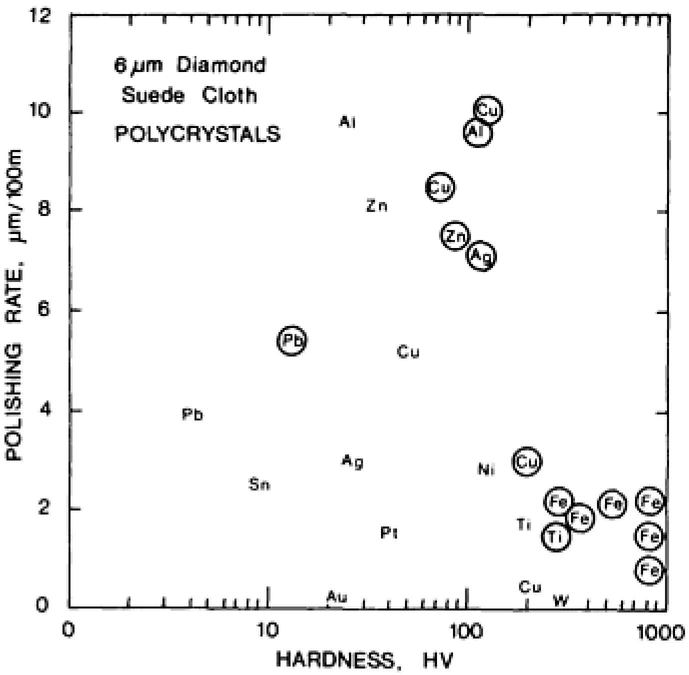


Fig. 10 Variation of polishing wear rate with workpiece hardness. Circled symbols indicate alloys of the base metal, whereas uncircled symbols indicate commercially pure metals. Wear rates were determined under comparable stringent conditions; the abrasive was under conditions that ensured the grits were held to the fibers of a polishing cloth by the method shown in Fig. 6(b). Source: Ref 15

The above considerations are concerned with the efficiency of the wear process. The absolute value of the wear rate depends also on the rate at which the surface grooves are formed; that is, the rate at which abrasive grits actually contact the wearing surface and become active. A very large quantity of grits is usually present in the general vicinity of a wearing surface, but very few of the particles are actually in contact with the workpiece surface. The fraction of contacting points will vary considerably with different types of systems, and this fraction is difficult to determine. Nevertheless, some qualitative indication of the importance of the factor can be obtained. For example, the wear rates obtained in deliberate polishing operations on cloths charged with an abrasive such that a comparatively high proportion of grits contact the polishing surface (as in Fig. 6b) are an order of magnitude higher than those obtained when few grits contact the polishing surface (as in Fig. 6a), even though the grits available per unit surface area of cloth are the same (Ref 15).

The fraction of contacting points available also contributes to the effect of abrasive particle size on the polishing wear rate because, for equal weights, the number per unit area of grits decreases rapidly with the increase in grit diameter. Figure 11 shows that the wear rate decreases markedly when the particle size is >3 μm (>120 μin.). This condition can be at least partially attributed to a decrease in the number of grits present per unit area of the polishing cloth. In this instance probable contributory cause is the fact that the fraction of the grits held in a workpiece-contacting position has also changed because it becomes more difficult to hold grits in a contacting position when the grit diameter approaches or exceeds the diameter of the cloth fibers (Fig. 5c). Under these conditions, the rapid decrease in wear rate when the grit diameter falls to <3 μm (<120 μin.) (Fig. 11) might seem to be paradoxical. However, the decrease in wear rate occurred and added a further complication because the predominant cutting mechanism began to be phased out, leaving only the inefficient delamination mechanism to provide the material removal process (see the section "Mechanisms of Material Removal" in this article).

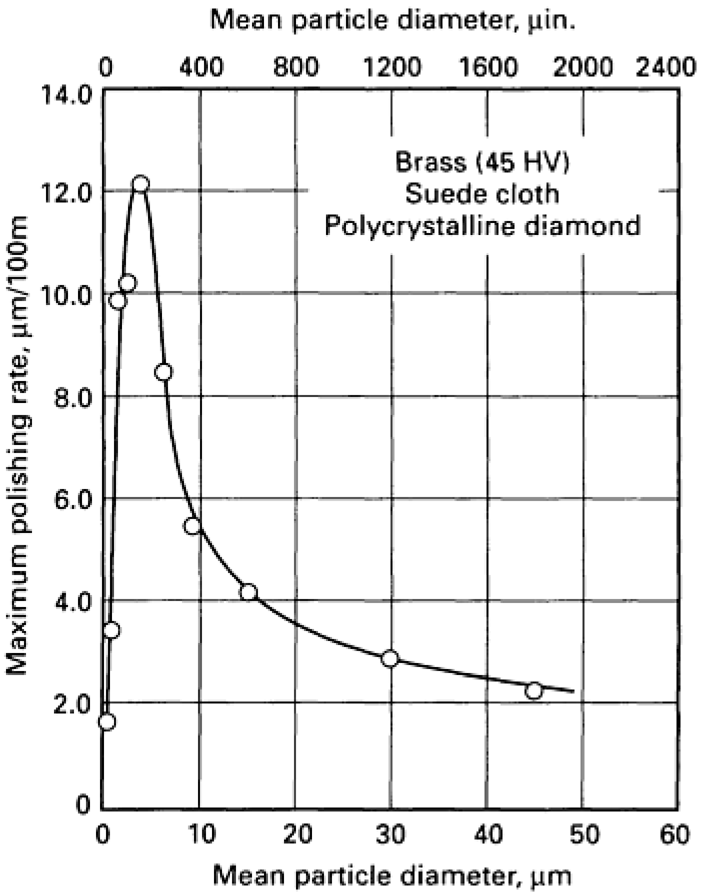


Fig. 11 Variation of the polishing wear rate with the particle size of a diamond abrasive used to machine 30% Zn brass

The polishing systems used in Fig. 10 and Fig. 11 can be regarded only as highly idealized and simplified models of wear systems that are typically encountered in engineering practice. These models certainly indicate, however, that the diagnosis and treatment of practical systems should be logically determined using a step-by-step procedure. The mechanisms of material removal certainly have to be established before rational analysis can commence. Subsequent comparisons with more idealized systems may indicate which methods of controlling the wear rate are feasible.

Polishing Wear without Abrasives

There are several possible mechanisms of polishing wear that do not depend primarily on the physical removal of material by the machining or scraping action of abrasives. Two such methods are discussed in this section:

- Surface flow
- Chemical-mechanical mechanisms

Surface Flow. It was once hypothesized that polishing was a thermally activated process in which a small volume at the crest of an asperity on a solid is heated to a high temperature for a brief period when an asperity on another solid contacts it and rubs past it. It was originally believed that the flash temperature reached the melting point of the solid of the pair that had the lower melting point and that some portion of this solid melted (Ref Ref 26, 27). It was then proposed that the molten material is transported from the asperity crest into an adjoining trough, thus gradually leveling the surface. The realization that the flash temperature might not always quite reach the melting point again altered the proposal for this process to suggest this crest material would be transported by plastic shear at high temperature. In either case, however, it was anticipated that the surface would gradually be made smooth and eventually become specularly reflecting (Ref 26, 27). In the strictest sense, the theory implies that small changes would occur in the dimensions of the surface being

polished but that no material would be removed unless the process were supplemented by a more conventional wear process. (This process would necessarily have to be one that does not destroy the specular reflectivity.) This theory became known as the Beilby theory of polishing; however, there is no need to elaborate on it further because conclusive evidence is now available that confirms it is invalid (Ref 15).

Nevertheless, it is still possible that the asperities on a moderately rough finished surface could be flattened by compression when rubbed by another stronger solid. Even modest localized heating of the workpiece surface during rubbing would increase this possibility. A specularly reflecting finish might then be obtained if the surface finish on the rubbing solid was conformed to specifications. This technique has been applied by jewelers, for example, to burnish precious metals by using tools of a polished hard material such as agate. However, wear in terms of material removal would again occur only if a subsidiary material removal mechanism did not destroy specular reflectivity during the machining process.

Chemical-Mechanical Mechanisms. Valid, if indirect, evidence exists that polishing can generate mechanisms which involve both chemical and mechanical elements in a symbiotic relationship (Ref 15, 28). The general concept is that the corrosion-protective film that normally covers a metal surface is removed from asperities by the rubbing solid. The protective layer is then reformed, consuming a thin layer of workpiece in the process. Repetition of the process gradually removes the crests of surface asperities, and the workpiece surface is leveled (Fig. 12). Only a small wear rate is expected, but a high polish is typically produced.

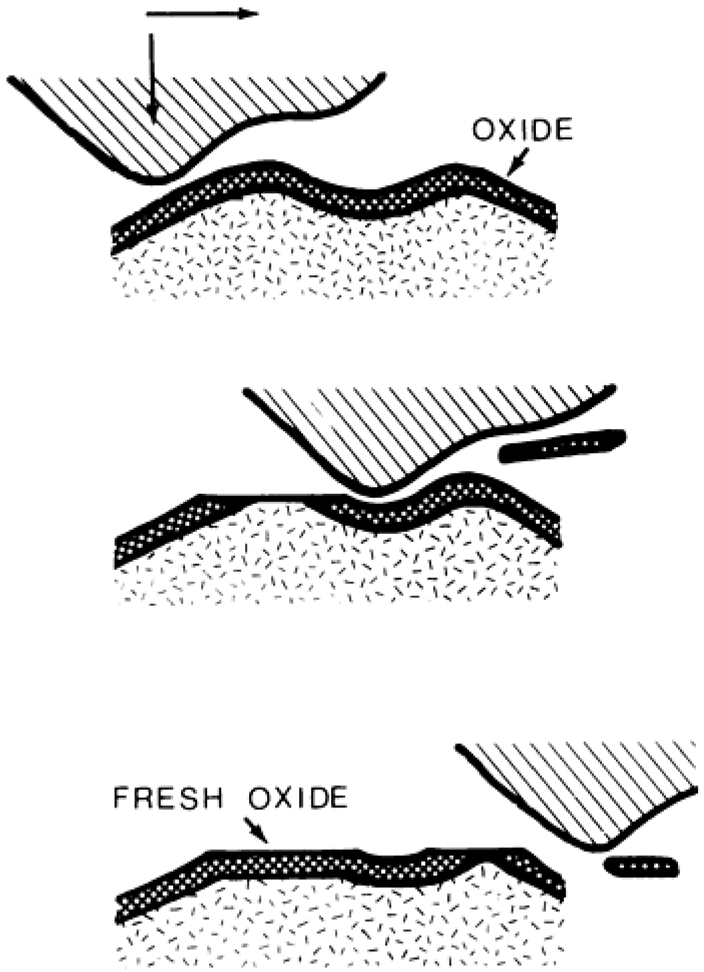


Fig. 12 Schematic of a proposed chemical-mechanical mechanism for the polishing wear process. Repeated removal and reformation of the corrosion-protective layer typically present on metal surfaces causes gradual leveling of the surface. The protective layer is shown to be an oxide and the interacting solid to be an abrasive grit; however, an asperity of any type would be equally effective provided that it was strong enough to break away the protective layer. Source: Ref 29

The rubbing should be, and indeed usually has to be, of a very gentle nature, but the contacting solid can consist of a range of types, including those with well-rounded contacting asperities. For example, rolling abrasives could be effective as could abrasives with well-rounded grits. This is in contrast to mechanisms that depend on abrasive machining. The presence of a medium that is only mildly corrosive would also be adequate. The process could theoretically occur, for example, in air because oxidation takes place the moment the protective film is removed. Nevertheless, the process is more likely to occur when a liquid (water being adequate in most cases) is present. More reactive liquids can, of course, accelerate the process. It is anticipated that metals whose apparent corrosion resistance depends on being covered by a self-healing surface film (that is, the more electronegative metals) would be most responsive to polishing wear by this mechanism, and that truly noble metals would be the least responsive. However, the production of a polish is in all cases dependent on a balance being achieved between the chemical and mechanical elements of the process. Even then, the development of some differences in the level of grains and constituents of the microstructure of the workpiece material is likely.

References

1. L.H. Tanner and M. Fahoum, *Wear,* Vol 36, 1976, p 299
2. R.S. Kirk, private communication
3. D. Tabor, *The Hardness of Metals,* Clarendon, Oxford, 1951
4. R.C.D. Richardson, *Wear,* Vol 10, 1976, p 291
5. L.E. Samuels, *Microindentation Techniques in Materials Science and Engineering,* P.J. Blau and B.R. Lawn, Ed., STP 889, American Society for Testing and Materials, 1986, p 5
6. R.M. Westrich, *Microindentation Techniques in Materials Science and Engineering,* P.J. Blau and B.R. Lawn, Ed., STP 889, American Society for Testing and Materials, 1986, p 196
7. J.B. Pethica and D. Tabor, *Surf. Science,* Vol 89, 1979, p 182
8. T.H. Kosel, *Microindentation Techniques in Materials Science and Engineering,* P.J. Blau and B.R. Lawn, Ed., STP 889, American Society for Testing and Materials, 1986, p 227
9. P.W. Bridgman, *Studies in Large Plastic Flow and Fracture,* McGraw-Hill, 1952
10. F.P. Bowden and T.P. Hughes, *Proc. Roy. Soc.,* Vol 160A, 1937, p 575
11. E.C. Rabinowicz, *Sci. Am.,* Vol 218 (No. 6), 1968, p 91
12. T.O. Mulhearn, *J. Mech. Phys. Solids,* Vol 7, 1959, p 85
13. A.J. Sedriks and T.O. Mulhearn, *Wear,* Vol 6, 1963, p 457
14. S.K. Dean, *The Production Engineer,* 1966, p 1
15. L.E. Samuels, *Metallographic Polishing by Mechanical Methods,* 3rd ed., American Society for Metals, 1982
16. L.E. Samuels and B. Wallace, *Metallography,* Vol 17, 1984, p 19
17. J.H. Richardson and R.A. Rose, *Microstructural Science,* Vol 4, 1976, p 281
18. T.O. Mulhearn and L.E. Samuels, *Wear,* Vol 5, 1962, p 476
19. E.D. Doyle and L.E. Samuels, *Proceedings of the International Conference on Production Engineering,* Japan Society for Precision Engineering, 1974, p 45
20. R.L. Aghan and R. McPherson, *Am. Ceram. Soc.,* Vol 56 (No. 1), 1973, p 46
21. B.R. Lawn and M.V. Swain, *J. Mater. Sci.,* Vol 10, 1975, p 113
22. E.D. Doyle and D. Turley, *Wear,* Vol 51, 1978, p 269
23. N.P. Suh, *Fundamentals of Friction and Wear of Materials,* D.A. Rigney, Ed., American Society for Metals, 1981, p 43
24. H.R. Shetty, T.H. Kosel, and N.F. Fiore, *Wear,* Vol 80, 1982, p 33
25. S.V. Prasad and T.H. Kosel, *Wear,* Vol 92, 1983, p 253
26. G. Beilby, *Aggregation and Flow of Solids,* Macmillan Publishing, London, 1921
27. F.P. Bowden and D. Tabor, *The Friction and Lubrication of Solids,* Clarendon, Oxford, 1950
28. E.C. Rabinowicz, *Friction and Wear of Materials,* John Wiley & Sons, 1965

29. T.F. Quinn, *Brit. J. Appl. Phys.*, Vol 13, 1962, p 33

Solid Particle Erosion

Thomas H. Kosel, Department of Electrical Engineering, University of Notre Dame

Introduction

SOLID PARTICLE EROSION (SPE) is the loss of material that results from repeated impact of small, solid particles. In some cases SPE is a useful phenomenon, as in sandblasting and high-speed abrasive waterjet cutting, but it is a serious problem in many engineering systems, including steam and jet turbines, pipelines and valves carrying particulate matter, and fluidized bed combustion (FBC) systems. Readers faced with problems in particular systems or materials should consult other articles in this Volume that deal with specific components or alloys.

Various design solutions have been developed in which high erosion rates are avoided by reconfiguring the system--such as the blocked tee configuration, in which a tee joint with one end closed is used in place of a gradual bend in a pipeline to prevent low-angle impingement. A good example of the variety of engineering solutions to SPE is provided by the case of power-generating steam turbines, in which exfoliation of iron oxide scale formed on steel heater tubes generates large pieces of scale that are fragmented into approximately 100 μm particles, causing erosion of turbine blades, shrouds, valves, rivets, and other components (Ref 1, 2). Liquid droplet erosion is also present. SPE solutions include minimizing of scale formation by using austenitic steels or chromizing treatments, particle removal with cyclones or screens, application of plasma-sprayed or diffusion coatings to blades, and redesign of turbine configurations.

This article is primarily a review of important SPE variables, their effects for different classes of materials, and the mechanisms and theories proposed to explain SPE. Because there is not general agreement in the literature on every point of practical importance, the reader will be alerted to areas that are less completely understood.

Solid particle erosion is to be expected whenever hard particles are entrained in a gas or liquid medium impinging on a solid at any significant velocity (greater than 1 m/s, or 3.3 ft/s). Manifestations of SPE in service usually include thinning of components, a macroscopic scooping appearance following the gas/particle flow field, surface roughening (ranging from polishing to severe roughening, depending on particle size and velocity), lack of the directional grooving characteristic of abrasion, and, in some but not all cases, the formation of ripple patterns on metals.

Solid particle erosion can occur in a gaseous or liquid medium containing solid particles. In both cases, particles can be accelerated or decelerated, and their directions of motion can be changed by the fluid. This is more significant in liquid media, and slurry erosion is generally treated as a different, though related, subject. In gaseous media, at least for particles larger than about 50 μm, deflection of the particles by the gas stream can often be ignored in erosion tests. However, it should be borne in mind that in an engineering application, such effects can be quite important--as, for example, in the extensive work of Tabakoff (Ref 3) on the spatial distribution and severity of erosion damage within turbines and the effects of particle size and the change of direction and speed when particles rebound from surfaces.

The distinction between erosion and abrasion should be clarified, because the term erosion has often been used in connection with situations that might be better classed as abrasion. Solid particle erosion refers to a series of particles striking and rebounding from the surface, while abrasion results from the sliding of abrasive particles across a surface under the action of an externally applied force. The clearest distinction is that, in erosion, the force exerted by the particles on the material is due to their deceleration, while in abrasion it is externally applied and approximately constant. This serves as a good working definition of the difference between the two phenomena. A clear-cut distinction between erosion and abrasion is difficult in some cases, particularly for very dense particle distributions in liquid or gas media, in which a "pack" of particles can develop and slide across the surface, which would be classed as abrasion according to our distinction. The remainder of this article will concentrate on SPE in dilute particle flows.

In practice, erosion rarely takes place in inert atmosphere or vacuum, but at room temperature the effect of oxidation can generally be ignored, and erosion thought of as "pure." Erosion/corrosion synergistic interactions are considered in the

final section of this article. Following Finnie (Ref 4), variables affecting pure erosion can be broadly separated into three types: impingement variables describing the particle flow, particle variables, and material variables. The primary impingement variables are particle velocity (*v*), angle of incidence (α), and flux (particle concentration). α is defined here and throughout virtually all erosion literature as the angle between the incident particle direction and the particle surface. Particle rotational speed is an additional impingement variable; it is difficult to measure and has not been studied in much detail. Particle variables include particle shape, size, hardness, and friability (ease of fracture). Material variables include all the material properties, such as hardness, work hardening behavior, and microstructure.

In the erosion literature, materials are broadly classified as ductile or brittle, based on the dependence of their erosion rate on α. Ductile materials, such as pure metals, have a maximum erosion rate, *E,* at low angles of incidence (typically 15° to 30°), while for brittle materials, such as ceramics, the maximum is at or near 90°. These two classical extremes are illustrated in Fig. 1. A variety of curves intermediate between these classical extremes exist, as discussed later, and in some cases the same material exhibits behavior that shifts from one extreme to the other, depending on erosion conditions.

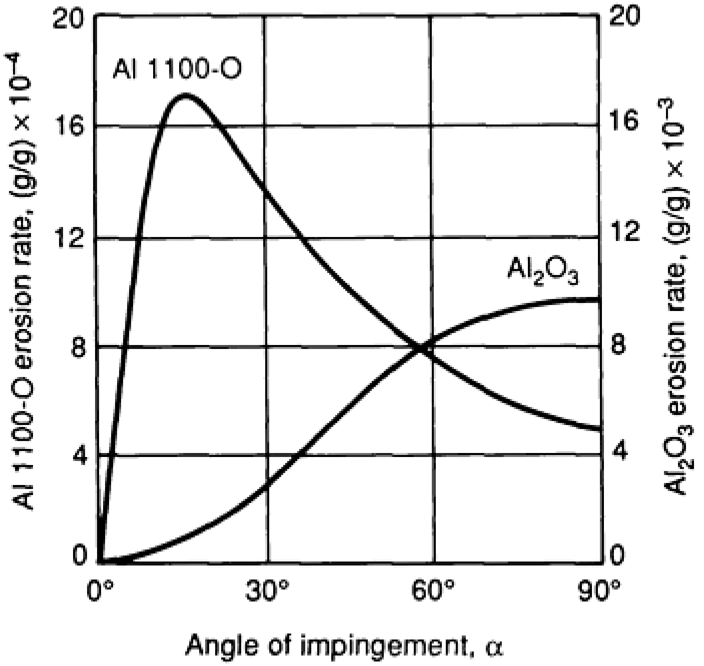


Fig. 1 Erosion of 1100-O aluminum relative to Al_2O_3 when both are eroded by 127 μm SiC particles impinging at a velocity of 152 m/s (499 ft/s). Source: Ref 4

Erosion

The phenomenology of erosion in the absence of corrosion will be briefly reviewed. The erosion rate, *E,* is commonly given in terms of mass or volume of material removed per unit mass of erodent impacted, volume being preferred because it permits thickness loss comparisons between materials of different density. Implicit is the assumption that the dimensions of the eroded area and the particle concentration are unimportant, which is a good approximation for dilute flows. Metals and ceramics differ in the dependence of *E* on α, as mentioned above, and also in their response to velocity and particle size and shape. *E* generally shows a power-law velocity dependence:

$$E = kv^n \quad \textbf{(Eq 1)}$$

where k is a constant and *n* is a velocity exponent that generally depends on material and erosion conditions. The value of *n* usually falls in the range of 2 to 2.5 for metals and 2.5 to 3 for ceramics, although observations exist outside these ranges. Material removal in ductile materials involves considerable plastic flow, while in brittle materials fracture is of primary importance, at least for higher angles of incidence. Theories predict that *E* should be inversely proportional to hardness for metals, while for ceramics there should be a much weaker dependence on hardness, but fracture toughness

should be important. Most theories further predict no effect of particle size for metals, although it is often observed that *E* increases strongly with particle size, at least up to about 100 μm. For ceramics, a particle size exponent is predicted and observed.

Mechanisms of material removal will be considered in more detail later, but it is evident that a division into metals and ceramics is justified. Because of the technological importance of steels and the failure of heat treatments that increase hardness to provide expected improvements in their erosion resistance, erosion of steels is treated in a separate section. A separate section on erosion of alloys containing large, hard second-phase particles, such as Stellites and high-chromium white cast irons, is also included. Such alloys provide excellent abrasion resistance, but under most erosion conditions they provide little or no improvement over plain carbon steels or even pure metals. The conditions under which such materials can provide significantly improved erosion resistance are considered.

Hansen (Ref 5) compared the erosion rates of a large number of alloys, ceramics, and cermets. Although limited to a particular set of conditions, Hansen's study provides the practicing engineer with a good comparison of erosion rates for a very broad range of materials under the same conditions. Estimation of the effects of variables such as velocity on the erosion rates of different classes of materials can be made from the remainder of this article. It should be noted that the order of material rankings would change with any change of variables such as velocity, particle type or size, and angle of incidence. The erosion tests were performed using 27 μm Al_2O_3 particles at normal incidence and 170 m/s (560 ft/s) at 20 and 700 °C (68 and 1300 °F) in nitrogen. As in most other studies, the tests were conducted with a gas-jet erosion apparatus in which particles are fed from a hopper into a nozzle, where they mix with and are accelerated by a flowing gas stream. ASTM has developed a standard test of this type (Ref 6), commonly used in industrial test labs. Hansen (Ref 5) normalized the erosion rates by defining the relative erosion factor (REF) as specimen volume loss divided by that of a standard material, Stellite 6B (Table 1). Figures 2 and 3 show his data for metals and ceramics, respectively. WC-Co cermets gave REFs from about 0.1 to 1.6, and REF was found to increase with binder content. The REFs of most metals were similar at 20 and 700 °C (68 and 1300 °F), typically within about 20% of unity (Fig. 2). The three lowest room-temperature REFs for metals were for tungsten (0.48), molybdenum (0.52), and 1015 steel (0.76), and the highest was for Ti-6Al-4V (1.26). The 700 °C (1300 °F) erosion rate of the standard (Stellite 6B) was 20% higher than the room-temperature value, so that 700 °C (1300 °F) REF values greater than 0.8 represent increases of erosion rate with temperature for a given material.

Table 1 Erosion test results for selected metals and ceramics evaluated at room temperature and at elevated temperature

Test parameters: 90° impingement; 27 μm Al_2O_3 particles; 5 g/min (0.18 oz/min) particle flow; 170 m/s (560 ft/s) particle velocity; 3 min test duration; N_2 atmosphere

Material	Manufacturing method[(a)]	Composition (manufacturer)[(b)]	Relative erosion factor (REF)[(c)]	
			20 °C (70 °F)	**700 °C (1290 °F)**
Metals				
Ti-6Al-4V	W	. . .	1.26	0.54
Haynes 93	C	17Cr-16Mo-6.3Co-3C-bal Fe (Stellite)	1.25	1.00
25Cr iron	C	25Cr-2Ni-2Mn-0.5Si-3.5C-bal Fe (OGC)	1.19	1.16
Stellite 6K	W	30Cr-4.5W-1.5Mo-1.7C-bal Co (Stellite)	1.08	1.06
Stellite 3	C	31Cr-12.5W-2.4C-bal Co	1.04	1.61
Stellite 6B	W	30Cr-4.5W-1.5Mo-1.2C-bal Co (Stellite)	1.00	1.00
Type 304 stainless steel	W	17Cr-9Ni-2Mn-1Si-bal Fe	1.00	0.73
Type 316 stainless steel	W	17Cr-12Ni-2Mn-1Si-2.5Mo-bal Fe	0.99	0.56
Haynes 188	W	22Cr-14.5W-22Ni-0.15C-bal Co (Stellite)	0.97	0.83
Haynes 25	W	22Cr-15W-10Ni-1.5Mn-0.15C-bal Co (Stellite)	0.96	0.85
Type 430 stainless steel	W	17Cr-1Mn-1Si-0.1C-bal Fe	0.93	0.62
HK-40	C	26Cr-20Ni-0.4C-bal Fe	0.93	0.78
Inconel 600	W	76Ni-15.5Cr-8Fe (HA)	0.92	0.61
RA 330	W	19Cr-35Ni-1.5Mn-1.3Si-bal Fe (RA)	0.91	0.79
Incoloy 800H	W	32.5Ni-21Cr-0.07C-46Fe (HA)	0.91	0.54
Beta III Ti	W	11.5Mo-6Zr-4.5Sn-bal Ti	0.90	0.57
Incoloy 800	W	32.5Ni-46Fe-21Cr (HA)	0.83	0.57
RA 333	W	25Cr-1.5Mn-1.3Si-3Co-3Mo-3W-18Fe-bal Ni (RA)	0.80	0.80
Inconel 671	W	50Ni-48Cr-0.4Ti (HA)	0.77	0.62

Mild steel	W	0.15C-bal Fe	0.76	. . .
Molybdenum	W	. . .	0.52	. . .
Tungsten	W	(GE)	0.48	0.17
Ceramics				
ZRBSC-M	HP	ZrB_2-SiC-graphite (N)	6.36	>5.00
Chromite	PS	(UCAR)	2.44	3.43
Refrax 20C	PS	SiC-Si_3N_4 bond (Carbor)	0.91	1.15
HD 435	. . .	Recrystallized SiC (N)	0.80	0.32
Carbofrax D	PS	SiC-ceramic bond (Carbor)	0.49	1.38
HD 430	. . .	Recrystallized SiC (N)	0.40	0.38
Si_3N_4	HP	(N)	0.40	0.12
Norbide	HP	B_4C (N)	0.38	0.21
BT-9	PS	2MgO-25TiB_2-3.5WC-bal Al_2O_3 (OGC)	0.37	0.36
BT-12	PS	1.5MgO-49TiB_2-3.5WC-bal Al_2O_3 (OGC)	0.35	0.16
BT-11	PS	1.7MgO-38TiB_2-3.5WC-bal Al_2O_3 (OGC)	0.33	0.26
ZRBSC-D	HP	ZrB_2-SiC (N)	0.32	0.07
BT-24	PS	2MgO-30TiB_2-3.5WC-bal Al_2O_3 (OGC)	0.32	0.20
BT-10	PS	2MgO-30TiB_2-3.5WC-bal Al_2O_3 (OGC)	0.30	0.25
Noroc 33	HP	Si_3N_4-SiC (N)	0.20	0.42
TiC-Al_2O_3	PS	(BW)	0.19	0.30
SiC	HP	(N)	0.12	0.02
CBN	. . .	(GE)	0	0
GE diamond	. . .	(GE)	0	0

(a) W, wrought; C, cast; HP, hot pressed; PS, pressed and sintered

(b) Manufacturer: BW, Babcock and Wilcox; Carbor, Carborundum Co., ; GE, General Electric Co; HA, Huntington Alloy Products; N, Norton Co.; OGC, Oregon Graduate Center; RA Rolled Alloys Corp; Stellite, Stellite Div., Cabot Corp.; UCAR, Union Carbide Corp.

(c) REF = Volume loss material/volume loss Stellite 6B

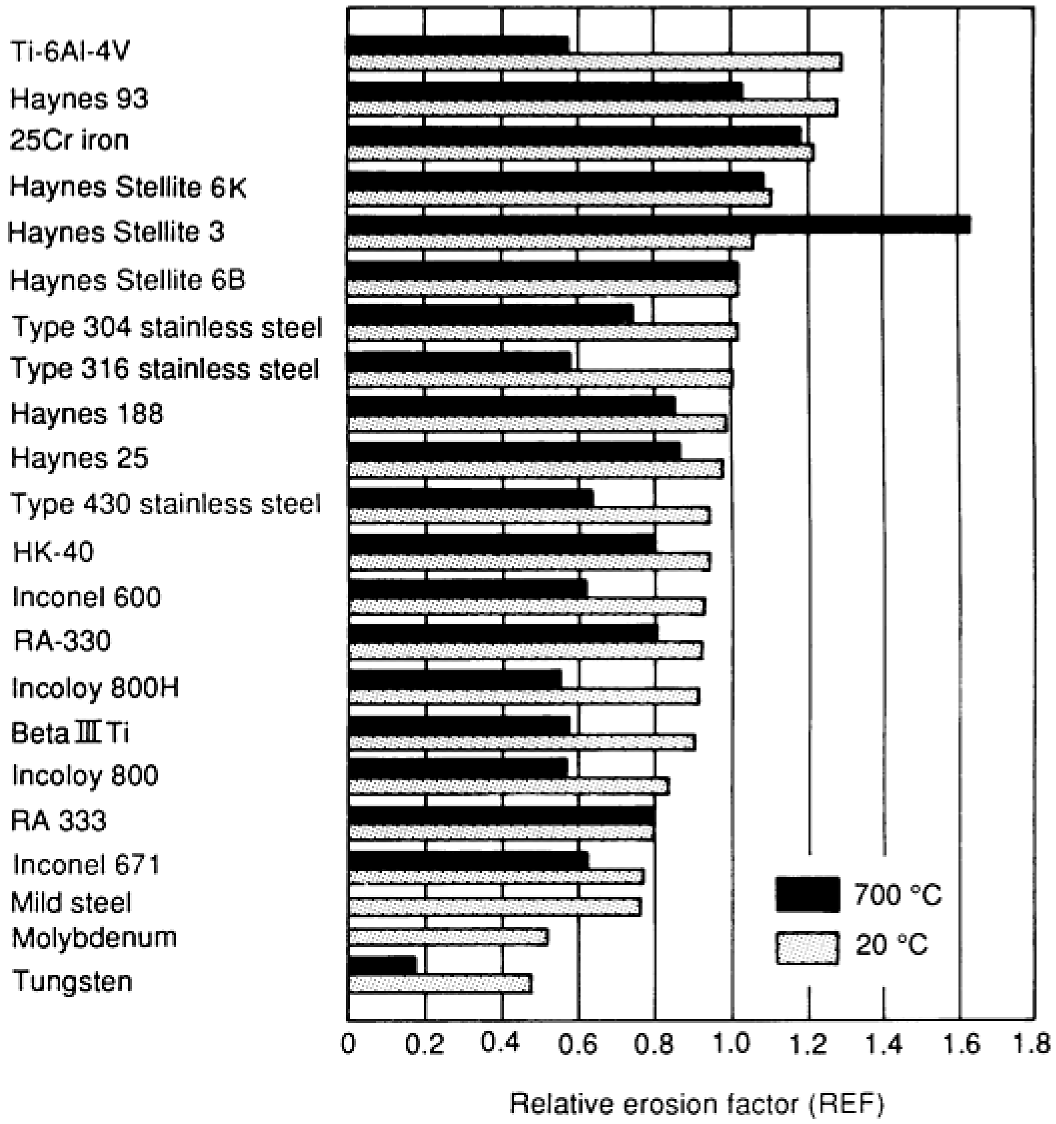


Fig. 2 Relative erosion factors for selected commercially available metals at an impingement angle of 90°. Stellite 6B cobalt-base alloy was used as the reference material. Source: Ref 5

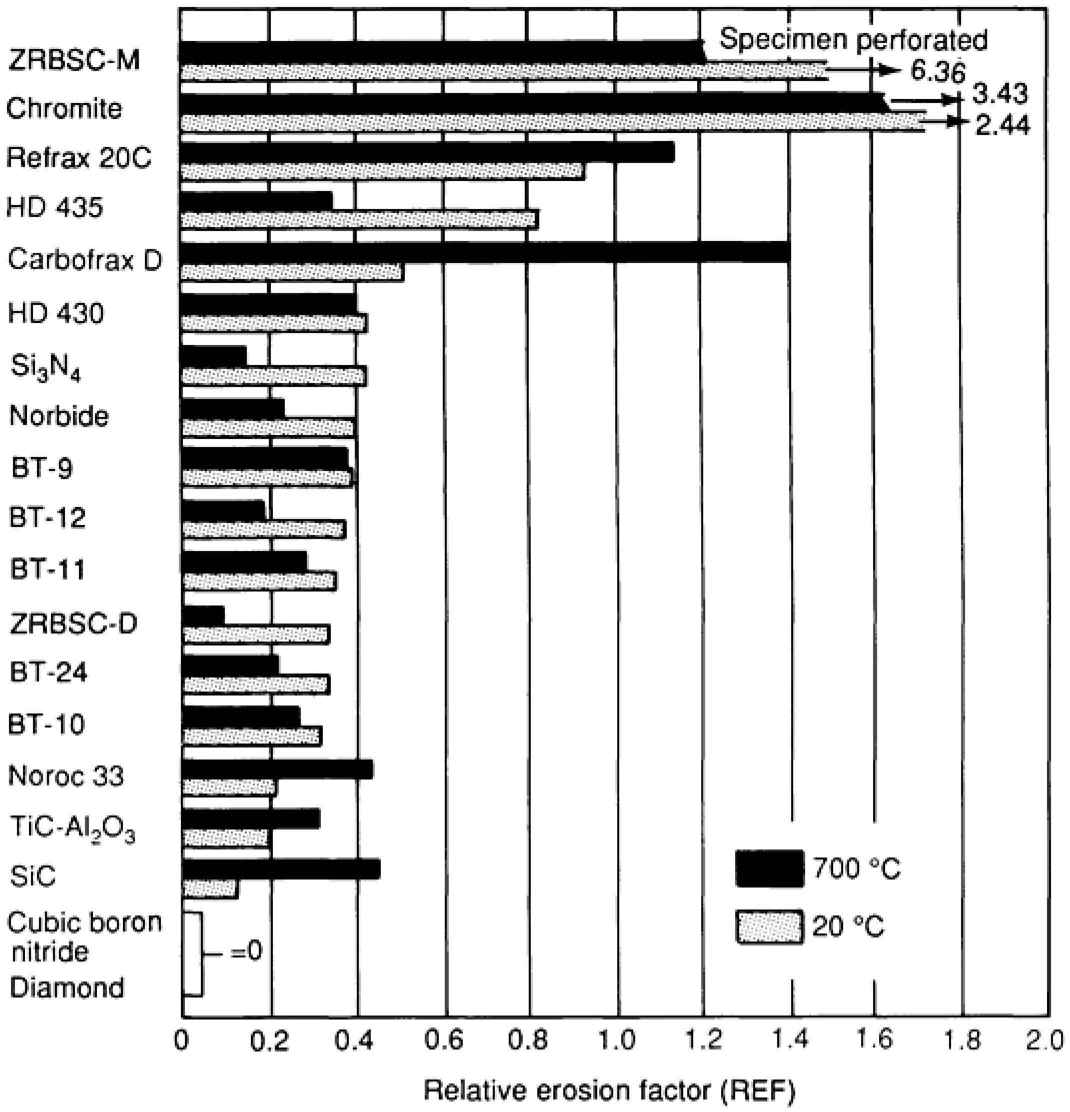


Fig. 3 Relative erosion factors for selected ceramics at an impingement angle of 90°. Ratings based on using Stellite 6B cobalt-base alloy as the reference material. Source: Ref 5

These results illustrate the unfortunate fact that alloy-strengthening mechanisms such as solution or precipitation hardening that increase hardness do not significantly improve erosion resistance. According to Hansen (Ref 5), if service experience reveals an erosion problem for a metallic component, substitution of another metallic alloy will generally provide little improvement. Most ceramics tested had REF values in the range 0.3 to 0.6, although a few were much higher, and a few were nearly zero. It is important to note here (as discussed later) that for erodent particles of lower hardness than Al_2O_3 (used in Hansen's study), significant improvements of erosion resistance can be obtained when the ratio of particle to target hardness, H_p/H_t, is less than 1.

Erosion of Metals

Throughout the large body of recent literature on erosion of metals are found the concepts of direct material removal by cutting or micromachining interaction of hard, sharp particles, and of repeated deformation of material leading to eventual removal of a flake or platelet or material. Direct material removal by cutting has been observed experimentally in some cases, as has the platelet mechanism, the latter has recently acquired a significant following. It is probable that both phenomena occur in varying proportions; further experimental and theoretical work is needed to clarify the importance of the two. From a design viewpoint, the distinction is important, because cutting models predict an inverse dependence of E on hardness, while ductility may be of greater importance for the platelet mechanism because it may affect the ease of

platelet removal. This section treats the erosion of single-phase metals and of metals for which the dimension of the impact damage is larger than the dimensions of second-phase particles.

The first theory for erosion of ductile materials was developed by Finnie (Ref 7, 8) and is still widely quoted. The trajectory of the particle was calculated using a number of simplifying assumptions, and the volume removed was equated to the volume swept out by the particle tip. The model predicted a velocity exponent of n = 2, and the predicted dependence of E on α was similar to experimental observations such as that for aluminum in Fig. 1, except for the fact that the model gave a zero erosion rate of 90° (normal) incidence. We will return to this point later. Finnie and McFadden (Ref 9) have discussed the fact that the velocity exponent often increases with increasing angle of impact, and have demonstrated that this can be accounted for by their microcutting analysis of low-angle impingement. E was predicted (Ref 7, 8) to be inversely proportional to the horizontal component of the flow pressure between the particle and surface. Together with the assumption that the particle did not rotate significantly during impact, this implies an inverse dependence on dynamic hardness, P--that is, the mean contact pressure during a high-velocity impact hardness test. Finnie (Ref 4) discussed the fact that P would not be expected to be equal to the static hardness, H, measured in standard hardness tests, because it involves high strain rates (typically 10^4 to 10^7), strains (several hundred percent), and possible temperature rise introduced by plastic deformation.

The observation by Finnie *et al.* (Ref 10) of an inverse dependence of E on H for pure face-centered cubic (fcc) metals and tantalum is illustrated in Fig. 4. The data are plotted as erosion resistance, defined as 1/E in units of g/mm^3, versus H. Finnie (Ref 4) noted that the deviation of points for body-centered (bcc) metals (iron, molybdenum, and tungsten) and steels was not surprising in view of the use of annealed static hardness values. In this regard, Finnie *et al.* (Ref 10) found that increasing H by cold working a metal has essentially no effect on E. It is commonly believed that this is because the surface reaches saturation work hardening during steady-state erosion. A more recent study by Goretta *et al.* (Ref 11) of the effect of cold rolling on erosion of pure copper, nickel, and type 304 stainless steel found that E changed by about 10 to 30% with cold work, but always by an amount significantly les than the change of H. Depending on the metal, α, and the amount of cold work, E either increased or decreased with H. From a practical point of view, cold work is clearly not a reliable method for improving crosion resistance.

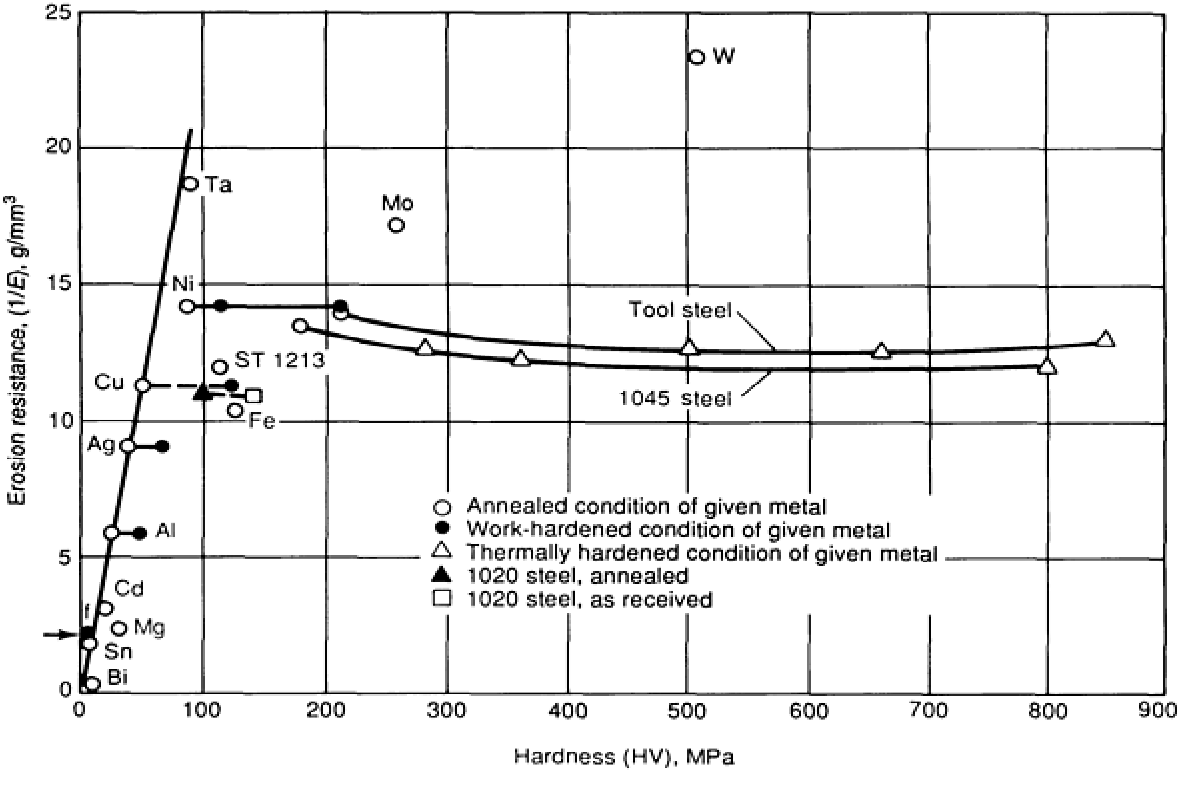


Fig. 4 Plot of erosion resistance versus static hardness for selected materials. Erosion was generated by 60-mesh or 250 μm SiC impinging at the surface at an angle of 20° with a velocity of 76 m/s (250 ft/s). Source: Ref 10

Mechanisms. Solid particles impacting metals form plastic impact craters and displace material. It has been shown that single impacts of spheres remove no material except above a rather high critical velocity (Ref 12) and that single, angular particle impacts remove material only for a small fraction of the possible particle orientations (Ref 13, 14). In the great majority of cases, direct material removal does not occur, but a lip of material remains attached at the end of the crater. Examples of the surfaces of pure nickel specimens eroded to steady state of 20° and 90° are included in Fig. 5; these are typical of the eroded surfaces of metals. Displaced lips of material are clearly visible at the ends of some impact creates on the surface eroded at 20°. As noted by Hutchings (Ref 15) and many other authors, including Finnie *et al.* (Ref 16), such lips are probably vulnerable to later direct removal or to being pushed back into the surface and becoming and extruded platelet that may eventually be removed by ductile fracture or other mechanisms (Ref 16). Finnie (Ref 7, 8) also noted that in order to account for the fact that the actual erosion rate is much lower than that predicted, it is necessary to assume that only a small fraction of the particles cut in the idealized manner, probably because additional impacts are needed to remove the material displaced to the end of the crater (Ref 16).

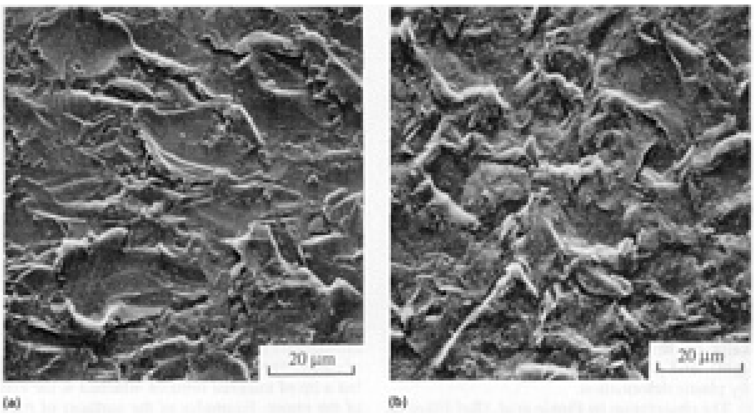


Fig. 5 Typical surfaces of ductile commercial-purity nickel eroded by 130 μm Al_2O_3 at two angles of incidence. (a) 20°. (b) 90°. Erodent velocity was 53.8 m/s (177 ft/s). Source: Ref 40

Finnie (Ref 17) described observations by Mayville (Ref 18) of the same surface area after successive impacts during normal impact erosion of copper and aluminum by SiC. Material removal typically involved the flattening of a protrusion into a platelike particle that was removed by one or two impacts. Typically, the volume removed per particle was smaller than the crater volume by a factor of about 200, emphasizing the multiple-impact nature of normal-impact material removal. Finnie *et al.* (Ref 16) described the mechanism of material removal for normal-incidence erosion of metals by angular particles as "a continuous 'battering' of the surface, leading to removal when extrusion of vulnerable material leads to a ductile fracture."

One of the most widely quoted studies on erosion mechanisms in metals is that of Bellman and Levy (Ref 19), who extended the sequential observation technique to include erosion at glancing angles of incidence. These authors provided convincing experimental evidence that one mechanism of material removal during erosion involves the deformation of surface material into thin platelets by repeated impacts. These eventually become detached and would be expected to form thin platelike debris particles. Figure 6 shows a series of micrographs of erosion of 7075-T6 aluminum at 30° with SiC at 30.5 m/s (100 ft/s) (Ref 19). The initial large crater in Fig. 6(a) is struck by another particle, forming a large platelet that covers the left of the crater in Fig. 6(b) and 6(c). The platelet is finally removed in Fig. 6(d). Platelets observed by Bellman and Levy were generally approximately equiaxed in the plane of the platelets. Many were considerably smaller than the average crater dimensions, although some were comparable to the crater size.

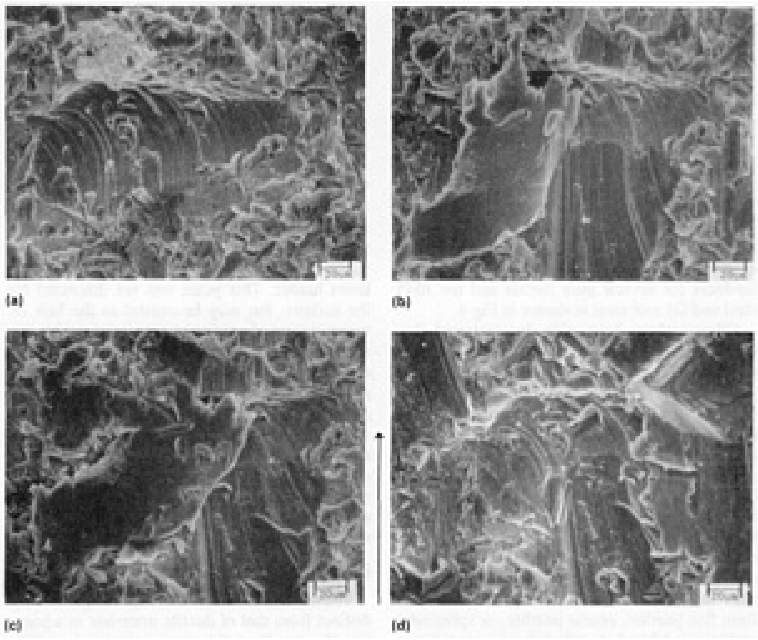


Fig. 6 Micrographs showing sequential erosion of 7075-T6 aluminum by SiC at a 30° angle of incidence with a particle velocity of 30.5 m/s (100 ft/s). The arrow indicates the direction of erodent particles projected onto the surface. See text for details. Source: Ref 19

Bellman and Levy (Ref 19, 20) believed that platelet formation is the primary and probably the only mechanism of material removal. They presented evidence (Ref 19) that part of an evaporated gold film was still present after a glancing impact, and Levy (Ref 20) showed that a glancing impact failed to remove a thin electroplated layer of copper on a steel specimen, suggesting that direct material removal by micromachining does not occur even at low angles of incidence. However, erosion debris particle studies (Ref 21, 22) indicate that micromachining is an operative mechanism at low α, although they do not permit a sufficiently accurate estimation of the relative importance of micromachining and platelet mechanisms.

The platelet mechanism has been summarized by Levy (Ref 23). One aspect of the model that has not been widely accepted is the assertion that surface temperature rise due to erosion leads to softening of the near-surface region, so that a layer of work-hardened material beneath this acts as an anvil against which the softer material is deformed. This aspect of the platelet mechanism was later considered by Hutchings and Levy (Ref 24), who estimated the temperature rise and considered published evidence for and against softening. They concluded that it was not necessary to postulate a thermally softened surface layer as an essential feature of the erosion mechanism under all conditions.

The platelet mechanism of erosion (Ref 23) has not been modeled quantitatively so as to make it possible to predict velocity exponents, the variation of E with α, H, or ductile fracture properties. The simple impact model of Finnie (Ref 7, 8, 9), which quantitative, also has shortcomings, particularly the prediction of zero erosion at 90° and the many observations that E is often not inversely proportional to H. Bitter (Ref 25, 26) offered a model that used the energy per unit volume required to remove material to characterize erosion. This model separated erosion into cutting and deformation wear, with the former peaking at about 20° incidence and the latter at 90°. The shapes of the E versus α curves for cutting and deformation modes are similar to those illustrated in Fig. 1 for ductile and brittle materials, respectively. Neilson and Gilchrist (Ref 27) used the same general approach to show that the E versus α curves of many materials could be fit by assuming the appropriate ratios of deformation and cutting energies per unit volume of material removed. Although these models suggest that different materials removal mechanisms predominate at 20° and 90°, neither theory offers clear mechanisms for either cutting or deformation wear. The platelet mechanism may be the predominant 90° removal mechanism for metals, and micromachining may predominate at low angles of incidence, but there is clearly not general agreement in the literature on these questions. Morrison *et al.* (Ref 28) argue that because velocity exponent, particle size effect, and surface features are similar for all angles of incidence, a single mechanism of erosion is appropriate at all angles of incidence. They consider that indentation, plowing, and cutting types of events, all of which involve similar plastic deformation behavior, occur at all angles, and the variation of the removal and nonremoval events with angle determines E versus α.

Particle Shape. Considerable work has been done on erosion by spherical particles, partly as an idealized case that is easier to analyze than angular particle impact and erosion. Computer models (Ref 12, 29, 30, 31) provide good agreement for rebound velocities and crater dimensions formed by single impacts. Steady-state erosion by spherical particles apparently generally develops a "hill and valley" surface topography, with the "mogul" spacing being significantly greater than single-impact dimensions. This is significantly different from the surface topography produced by angular particles. Brown *et al.* (Ref 32) found that material removal occurred both by plowing deformation on the sides of the hills and by removal of flakes of material from the valleys. Considerable embedding of glass was observed to play a role in material removal in the valleys, with a glass/metal composite layer being formed, and recrystallization also played a role.

As might be expected, angular particles give higher erosion rates than spheres. For example, Maji and Sheldon (Ref 33) observed increases of 70°, 20 m/s (65 ft/s) erosion of 6061-T6 aluminum by about a factor of 5 for 270 μm angular hardened steel grit compared with spherical shot of the same size, and also showed that fragmentation of the spherical shot caused a significant increased of *E*. The degree of angularity of angular particles is also important. Methods of quantifying angularity by means of image analysis, and is effect on *E*, have been discussed by Bahadur and Badruddin (Ref 34), who found that *E* increased by about a factor of 2 to 3 with angularity, which is found to vary with grit size.

Embedding of Erodent Fragments. The influence of embedding of fragments of the erosive particles in the specimen surface is in some cases considerable. Edington and Wright (Ref 35, 36) showed evidence of embedding of particle fragments in Stellite 6B to form a continuous surface layer. Similar observations were made by Ruff *et al.* (Ref 37, 38) for Al_2O_3 erosion of copper and by Kosel *et al.* (Ref 39, 40) for Al_2O_3 erosion of nickel. The work of both Edington and Wright (Ref 35, 36) and Kosel *et al.* (Ref 39, 40) showed that normal-incidence erosion produced a continuous composite surface layer, consisting of particle fragments mixed with the eroded metal, which is actually self supporting when the underlying substrate is electropolished away. The observations of Kosel *et al.* (Ref 39, 40) using Auger spectroscopy showed that the outermost surface contained only about 10% metal for nickel eroded by 400-grit Al_2O_3 at normal incidence and 53 m/s (175 ft/s). Observation by transmission electron microscopy (TEM) showed that the Al_2O_3 particle size was only on the order of 0.1 μm, compared with 23 μm for the incident particles, indicating that embedded fragments were subsequently refragmented by further impacts. 100-grit Al_2O_3 produced a discontinuous composite layer at normal incidence and 53 m/s (175 ft/s), but generated a continuous layer when the velocity was reduced to reduce the initial crater size. Both Kosel *et al.* (Ref 39, 40) and Ives and Ruff (Ref 37) described qualitative models for material removal under conditions giving rise to a continuous embedded layer in which incident particles never impinge directly on the substrate metal.

Particle Size. The size of erodent particles has little or no effect on the erosion rate of ductile materials as long as the particle size is above about 100 μm, but the erosion rate decreases rapidly with decreasing particle size below 100 μm. This fact, known as the size effect, is also found in abrasive wear. Misra and Finnie (Ref 41) discussed the many theories that have been suggested to explain this effect and concluded that it is probably due to an actual increase in the flow stress with decreasing indentation size. Evidence for this has been presented by Gane and Cox (Ref 42, 43), who performed microhardness indentation tests. Some recent studies have found that *E* is not always constant for particle sizes above approximately 100 μm. For instance, Liebhard and Levy (Ref 44) found that *E* became constant for SiC sizes above 200 μm at 20 m/s (65 ft/s) on 1018 steel, but continued to rise even up to 850 μm particle sizes for 60 m/s (195 ft/s) erosion of the same material.

Particle Hardness. The hardness of the erodent particles relative to the material being eroded is an important, often overlooked, factor in erosion. As in the case of abrasive wear (Ref 45, 46), it was found by Wellinger and Uetz (Ref 47) that the erosion rate drops dramatically when particle hardness decreases below that of the material being eroded. This and other similarities between erosive and abrasive wear have been discussed by Misra and Finnie (Ref 41). As noted by Finnie (Ref 17), it is quite possible that the observation that (as will be discussed later) heat treatment of steels appears to have almost no effect on their erosion resistance may be partially due to the use of hard particles such as SiC and Al_2O_3 in most previous work.

Particle Flux. Impact models predict no effect of erodent particle flux on erosion, because only one particle is considered. Anand *et al.* (Ref 48) have demonstrated that because of interference of rebounding particles with incident particles, the surface is effectively shielded--the result being that *E* decreases exponentially with increasing flux. For example, in the case of 1018 steel eroded with 63 μm Al_2O_3 at 90° and 50 m/s (165 ft/s), the erosion rate with a feed rate of 5 g/min (0.18 oz/min) in a 3.2 mm ($\frac{1}{8}$ in.) diam nozzle is reduced by 10% due to this effect. The effect increases with decreasing velocity or particle size and depends on material through the effect of rebound velocity, which affects the time during which the rebounding particles shield the surface.

Temperature. The influence of temperature on erosion (in the absence of complications caused by corrosion) is not as simple as might be expected on the basis of decreasing yield strength with increasing temperature. Sundararajan (Ref 49) has reviewed this area; in some cases *E* increases with temperature, while in others it decreases. Some data are included in Fig. 2 and 3.

Erosion of Steels

Erosion of steels is treated separately, because although the hardness of steels can be dramatically increased by heat treatment, there is no evidence that this improves erosion resistance. The accompanying decrease of tensile ductility might also be expected to cause a shift to more brittle behavior. Finnie (Ref 50) found that the erosion rate of 1055 steel by 250 μm SiC peaks at about 25° in the ferrite plus pearlite condition, but at 90° for the same steel in the martensitic condition, using 250 μm SiC at 107 m/s (350 ft/s). Other authors (Ref 51, 52) have found no effect of heat treatment on the shape of the *E* versus α curve for 36 μm Al_2O_3. This could possibly be due to the use of larger particles in Finnie's study (Ref 50), since Sheldon and Finnie (Ref 53) showed that the *E* versus α curve of a hardened tool steel shifted from a shape intermediate between brittle and ductile materials with approximately 125 μm SiC to a clearly ductile shape with a maximum at 20° with approximately 9 μm SiC. Similar behavior was also found for MgO and plate glass, and with 9 μm particles the surface ripple pattern characteristic of ductile metals (Ref 54) occurs at low α even for MgO and glass. The ductile behavior of brittle materials with small particles is probably directly connected to the fact that below a critical load and indentation depth, cracks are not formed in Vickers microhardness tests on ceramics. This effect is known to occur in abrasion as well. It is not clear whether the 30 μm Al_2O_3 used by Sargent *et al.* (Ref 51) and Peterson *et al.* (Ref 52) should be small enough to cause the transition to ductile behavior, but the latter authors searched for and found no evidence of brittle behavior of martensite.

The most important theoretical and technological aspect of erosion of steels is that several authors have demonstrated that the erosion rate at fixed α is nearly independent of hardness in individual steels, contrary to Finnie's theory (Ref 7, 8, 9). For example, the plot given by Finnie *et al.* (Ref 10) for 20° erosion resistance versus hardness for several pure metals and for 1045 steel and O1 tool steel is shown in Fig. 4.

Other authors have also demonstrated that hardness has little effect on erosion resistance in a given steel and that *E* sometimes increases slightly with *H.* The results of Sargent *et al.* (Ref 51) indicated very little effect of hardness at any impact angle. Salik and Buckley (Ref 55) found almost no effect of hardness on erosion for 1045 steel eroded by fine ($\sim$100 μm) crushed glass at 90° incidence. These authors did find a very substantial increase in erosion resistance at the highest hardness level, using 15 μm glass beads, but not with crushed glass. This may be an affect unique to spherical particles. Levy and Jahanmir (Ref 56) found that 1075 steel heat treated to form fine pearlite, coarse pearlite, or spheroidal carbides exhibited ductile behavior and that a slight increase in erosion occured with increasing hardness.

Zhu and Mao (Ref 57) studied erosion of steels and three WC cemented carbides by particles with hardnesses in the range of 90 to 2500 HV. A steel equivalent to 4140 was tested in the annealed condition (254 HV) and two quenched and tempered conditions (464 and 642 HV), and a high-speed tool steel was tested in the annealed (289 HV) and quenched (923 HV) conditions. In general, soft erodents produced maximum erosion rates at or near 90°, while erodents significantly harder than the target produced maximum erosion at low angles. The 4140 and tool steels showed only about 10% variation in *E* with hardness for SiC ($\sim$2500 HV), Al_2O_3 ($\sim$2200 HV), and SiO_2 ($\sim$1350 HV). However, with angular glass ($\sim$500 HV) erodent, the 15°, 65 m/s (210 ft/s) erosion rate of the 4140 steel decreased from about 0.41 to 0.25 mg/g due to hardening, and for the tool steel *E* decreased from about 0.28 to 0.08 mg/g upon hardening. The hard erodents produced typical features suggesting micromachining at low α, while the soft erodents produced surface deformation without clearly defined grooves.

Peterson *et al.* (Ref 52) studied erosion of 1078 and 10105 steels in spheroidized, pearlitic, tempered, and untempered martensitic conditions, with *H* ranging from 163 to 775 HV. Velocities from 40 to 100 m/s (130 to 330 ft/s) and angles from 15° to 90° were used, with 240-grit (36 μm) Al_2O_3. The peak erosion rate was about 40° at 40 m/s (130 ft/s), but increased to near 90° at 100 m/s (330 ft/s). *E* was essentially independent of *H* at 40° and increased somewhat with *H* at 90°. Surface features indicated ductile displacement of material for all microstructures, including untempered martensite, and no evidence of subsurface cracking was found in cross sections. There was remarkably little difference between micrographs of spheroidized and untempered martensitic 10105 steel at 90° incidence, either in general appearance or in mean crater dimensions. The latter is consistent with the observation of similar erosion rates (0.14 and 0.24 mg/g, respectively), but is clearly not consistent with impact theories, which would predict much smaller crater dimensions for the martensitic microstructure because it was 4.4 times harder. This point was not discussed by the authors, but may be central to the lack of improvement of erosion resistance of steels with *H.* Because there is no evidence of different

response in terms of brittle versus ductile fracture, the material removal rate would be expected to be essentially proportional to crater volume. Some unpublished work by Liang and Kosel (Ref 58) has shown that the dynamic hardness of 1080 steel does not increase as rapidly as static hardness, but that the difference is not sufficient to explain the lack of change of erosion rate.

Erosion of Ceramics

The erosion of brittle materials is significantly distinct from that of ductile materials in a number of ways. First, the erosion rate peaks at 90°, as indicated in Fig. 1. The mechanism of material removal is also markedly different, involving cracks initiated by brittle fracture, at least for normal-incidence erosion. The influence of erodent particle size on erosion is also different in brittle materials.

The erosion behavior of ceramic materials has been reviewed by Ruff and Wiederhorn (Ref 59), Wiederhorn (Ref 60), Evans (Ref 61), and, most recently, Routbort and Scattergood (Ref 62). A book on the subject is also in press (Ref 63). It is well established that erosion of ceramics involves formation and propagation of cracks to generate chips of material. Most authors quote the work of Lawn and Wilshaw (Ref 64), who showed that during static indentations, radial ("penny") cracks and/or conical (Hertzian) cracks form during loading, whereas lateral vent cracks form during unloading. Hertzian cone cracks, in which the cone vertex lies at or near the surface, are the only cracks formed by spherical or blunt particles at low impact velocities (Ref 59). Erosion of glass by spherical glass beads was studied by Mehrotra *et al.* (Ref 65), who found that material removal proceeds initially by interaction of cone cracks and later by chipping. Sharp or angular particles generally produce radial cracks, which lie perpendicular to the surface and are semicircular in shape, and which form during loading. Upon unloading, the stresses generated by the plastically compressed zone cause the formation of lateral vent cracks approximately parallel to the surface. These initiate beneath the indentation and travel upward to the surface, resulting in material removal from the surface. In contrast, radial cracks alone cannot result in material removal, although they are often intersected by subsurface lateral vents, giving rise to removal of pie-shaped wedges of material. Such behavior was observed in MgO, for instance, by Rickerby and Macmillan (Ref 29). Radial cracks, being approximately perpendicular to the surface, reduce the strength of the ceramic by acting as preexisting flaws. Figure 7 shows an example of a single impact crater on sapphire in which material removal by a large semicircular lateral vent crack is evident.

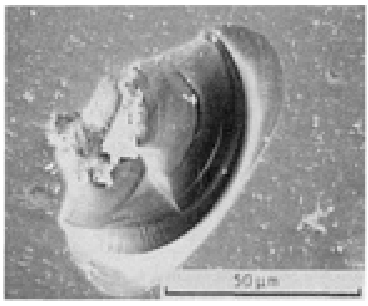


Fig. 7 Surface damage in sapphire (Al_2O_3) produced by 90° impact of 405 μm SiC particles traveling at a velocity of 90/ms (300 ft/s). Note material removed by the extension of a large semicircular lateral vent crack. Source: Ref 73

Considerable evidence for plastic deformation in nominally brittle ceramics has been presented by Hockey *et al.* (Ref 66), who used TEM to observe the dislocation structures beneath impact craters in MgO, silicon, SiC, and Al_2O_3. Such plasticity is probably due largely to the compressive component of the stress field below the impacting particle. High dislocation densities were observed in all cases, although the relative extents of the dislocation damage and crack propagation compared with the indentation size varied with the material.

At low angles of incidence, angular particles still produce high dislocation densities, and surface observations show that plastic displacement of material occurs in a similar manner to that in ductile materials (Ref 59). The extent of cracking is greatly reduced, and it may be that plastic deformation is of primary importance in erosion at glancing in incidence. In some instances, TEM observations (Ref 66) showed dislocation tangles without evidence of cracking for 15° incidence of single particles. Steady-state surfaces showed considerable evidence of plowing, with only minor cracking along the edges of the furrows. It is interesting to note in this context that Shetty *et al.* (Ref 67) observed the formation of

continuous micromachining chips from large M_7C_3 carbides in Stellite alloys during scratch tests with hard particles, again indicating the role of plasticity in wear of brittle phases.

Two major elastic-plastic theories of erosion have been developed. The theory by Evans *et al.* (Ref 61, 68) includes the effects of elastic waves generated by the impacting particle, as well as plastic waves. The theory of Wiederhorn *et al.* (Ref 59, 60, 69) assumes that dynamic effects may be neglected and that the particle penetration depth is determined by equating the plastic work done to the incident kinetic energy of the particle. The erosion rate predicted by Evans *et al.* is given by:

$$E \propto v_0^{3.2} d^{2/3} \rho^{1.3} K_{\mathrm{Ic}}^{-4.3} H^{-0.25} \quad \textbf{(Eq 2)}$$

while that predicted by Wiederhorn *et al.* is:

$$E \propto v_0^{2.4} d^{2/3} \rho^{1.2} K_{\mathrm{Ic}}^{-4.3} H^{0.11} \quad \textbf{(Eq 3)}$$

where v_0, d, and ρ are particle velocity, diameter, and density, respectively, H is material hardness, and K_{Ic} is material toughness (Ref 59, 60). It should be noted that the above expressions are valid when E is in units of mass or volume loss per unit mass impacted, as most experimental data are given in this form. The original theories were formulated in terms of material loss per impact, and because the mass per particle is proportional to d^3. the particle size dependence in these units is $d^{1\ 1/3}$ (Ref 62). The models can be used to predict E for $\alpha \neq 90°$ by using the normal component of velocity, $v \cdot \sin \alpha$ (Ref 66), and particle size and velocity thresholds can be included by replacing d by ($d - d_0$) and v by ($v - v_0$) (Ref 70).

Studies of single crystals or glass permit assessment of the theories without microstructural complications (Ref 62). Routbort *et al.* (Ref 70) studied erosion of silicon single crystals and found that, as shown in Fig. 8, E versus $v \cdot \sin \alpha$ gives the same value and velocity exponent except for low α. Similar results were observed for other ceramics in other studies. This effect was attributed to plasticity at low α, but Srinivasan and Scattergood (Ref 71) showed that it could be accounted for by greater crack extension at low α due to changes in stress distribution. Velocity exponent is found to decrease with increasing particle size, as indicated by the data points in Fig. 9. This is not explained by existing erosion models (Ref 62), whose predictions are indicated by the dotted lines, but may be due to greater temperature rise for larger particles (Ref 72) or to particle blunting and fragmentation (Ref 73, 74). Little work has been done to test the prediction that E is proportional to $d^{2/3}$, but work on silicon is in agreement with this (Ref 70).

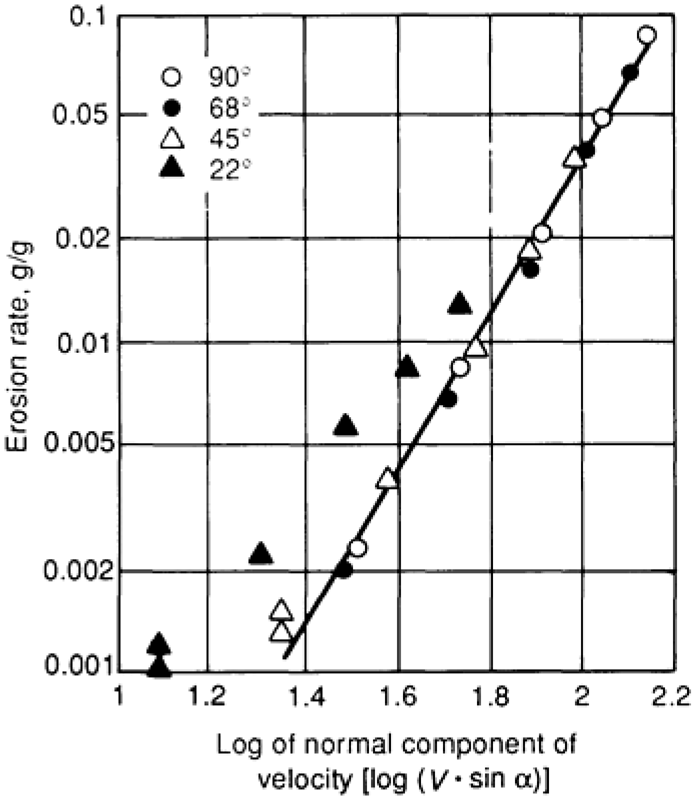


Fig. 8 Plot of erosion rate versus the normal component of velocity ($v \cdot \sin \alpha$) for a (111) silicon single crystal as a function of angles of incidence. Al_2O_3 erodent particle size was 390 μm. Source: Ref 62

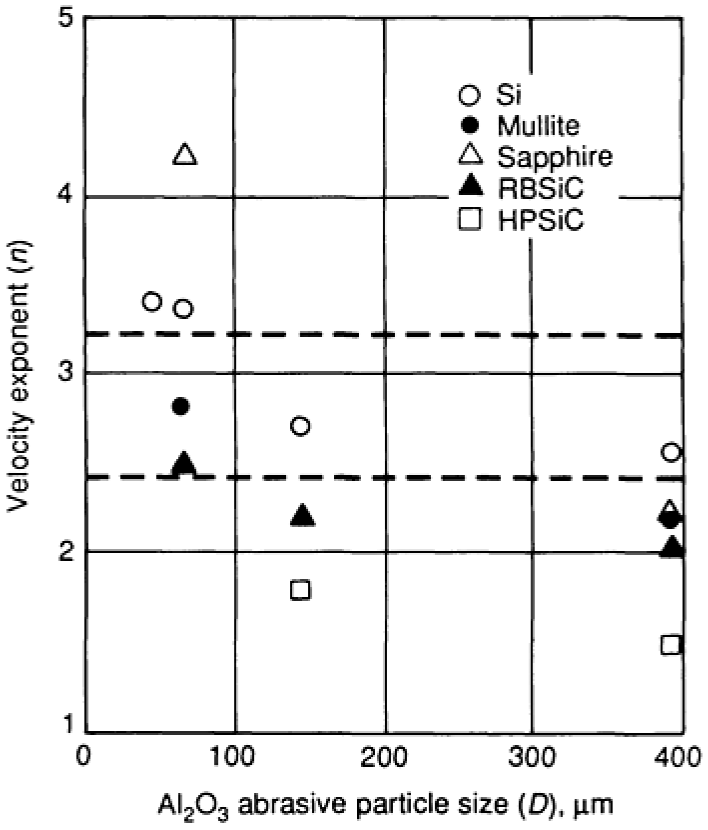


Fig. 9 Plot of velocity exponent versus Al_2O_3 particle size for selected ceramics. Source: Ref 62

Erosion of conventional ceramics (Al_2O_3, SiC, Si_3N_4) is also in general agreement with the models in most cases. The velocity exponent is lower for softer Al_2O_3 erodent than SiC because of the low particle-to-target hardness ratio; no cracks form in these ceramics when eroded by Al_2O_3 (Ref 75). Studies of Al_2O_3 by Ritter *et al.* (Ref 76, 77, 78, 79, 80) reveal that for large-grained material, grain ejection can occur with very irregular radial crack patterns. Although it might be expected that toughened partially stabilized zirconia (PSZ) would give improved erosion resistance, after an initial improvement E saturates and becomes nearly independent of K_{Ic} (Ref 81). A number of studies of ceramics toughened by addition of ceramic whiskers have been reviewed by Routbort and Scattergood (Ref 62). Improvements of erosion resistance have been obtained, but are not always in agreement with the $K_{Ic}^{-4/3}$ prediction of the models. Additional factors suggested as important in erosion of ceramics are (Ref 62) erodent fragmentation and *R*-curve effects (the dependence of crack growth resistance on crack size).

Particle Hardness. For quartz (~1100 HV), which is the hardest constituent of most minerals, the erosion rates of ceramics can be much lower than for Al_2O_3 (~2000 HV). Gulden (Ref 82, 83) studied erosion of reaction-bonded (RB) and hot-pressed (HP) Si_3N_4, glass-bonded Al_2O_3, and MgF_2. Hot-pressed Si_3N_4 was eroded by both SiC (~2500 HV) and quartz, and for a given combination of particle size and velocity, E was two orders of magnitude lower for the softer quartz erodent. Scanning electron microscopy (SEM) studies of single impacts also showed very small indentation for quartz, with none of the material removal by lateral fracture which occurred for SiC particles, so that the impact damage dimensions shown for comparable velocity and particle size was about 100 times smaller for quartz. Routbort *et al.* (Ref 84) compared erosion rates of three whisker-reinforced ceramic-matrix composites with matrices of Al_2O_3, Si_3N_4, and stabilized ZrO_2 by four erodents. Increases of E by factors of 10 to 20 were observed for SiC erodent (~2500 HV) compared with Al_2O_3/ZrO_2 (~1500 HV). In simple terms, erosion by particles softer than a ceramic target is expected to deform the particle as well as the target. Murugesh and Scattergood (Ref 73) recently studied erosion of several grades of Al_2O_3, including sapphire, by SiC and Al_2O_3. The harder erodent typically gave higher erosion r0ates by factors of only 2 to 4, and there was not a major increase of erosion rate for $H_p/H_t > 1$, although values much less than 1 were not investigated. The results were interpreted in terms of a model for erodent fragmentation, which can reduce E by particle blunting.

From a design standpoint, the important point is that particles encountered in practice are often significantly softer than ceramics, and if this is the case in a particular application, use of ceramics could be very beneficial if permitted by other design requirement such as toughness and conductivity. Ceramic coatings applied by various techniques can also offer significant advantages, provided they are thick enough to survive the service life of the component and to confine impact damage within the coating. Thermal expansion mismatch problems must also be avoided.

Erosion of Coarse Two-Phase Microstructure

This section will discuss erosion of alloys with coarse, nominally two-phase microstructures in which the second-phase particles (SPPs) are typically large compared with the dimensions of the damage zone created by the impact of one particle. Many other alloys that have multiple phases, such as steels and precipitation-hardened alloys, can be treated as though they were homogeneous, because the impact damage occurs over dimensions large compared with the microstructure.

Given the many similarities between erosion and abrasion (Ref 85), highly abrasion-resistant materials, such as white cast irons and the Stellite alloys, might be expected to exhibit high erosion resistance. Ideally, one might hope to combine the relatively good erosion resistance of a hard ceramic phase with the desirable ductility and toughness of a metal by choosing an alloy with a ductile metallic matrix and hard SPPs, such as carbides. However, as discussed recently by Kosel and Ahmed (Ref 86), this is rarely the case, due to the synergistic increase of the erosion rate of a hard, brittle phase by its presence as a dispersed phase in a relatively soft matrix.

Aptekar and Kosel (Ref 87) studied erosion of high-Cr/Mo white cast irons and Stellite alloys by angular Al_2O_3 and by both angular and rounded quartz particles. Because quartz is softer than the carbides in the alloys, while Al_2O_3 is of comparable hardness, quartz does not cut the carbides directly during abrasion (Ref 88, 89, 90), while Al_2O_3 does so quite readily, leading to a reduction of abrasion rate by a factor of about 25 for quartz compared with Al_2O_3 on Stellite. However, the reduction of erosion rate for quartz compared with Al_2O_3 was only about a factor of two at most (Ref 87). E also increased with second-phase volume fraction (SPVF) for both Al_2O_3 and crushed quartz, demonstrating that the SPPs

could actually be detrimental to erosion resistance. The erosion rate of white cast iron by crushed quartz was a higher than that of annealed 1020 steel, and the large primary carbides were deeply depressed below the surface.

The failure of large carbides to improve erosion resistance led to a study by Kosel and Aptekar (Ref 91) of erosion of model alloys consisting of Al_2O_3 or WC/W_2 CSPPs dispersed in a matrix of pure copper. *E* typically increased rather than decreased with SPVF, even when *E* was 20 to 80 times lower for a flat plate of the ceramic phase than pure copper. Variations of *E* and SPP protrusion with SPP size and with erodent particle velocity and size were consistent with the edge-effect concept described more fully below.

The edge effects refers to the increase of *E* for a ceramic due to easy fracture at SPP edges. If a large, flat plate of ceramic is eroded, its edges become rounded to a steady-state width, *W,* from the original edge, but the removal rate of the interior of the plate is unaffected. However, if the width of the plate is less than 2*W,* the entire surface of the plate is subject to easy chipping, and erosion rate increases over that for an infinite plate by an amount that increases with decreasing plate width or increasing *W.* Figure 10 illustrates a narrow plate of width *d,* or a rod of diameter d_r, eroded from the top at α= 90°. The two curves show the superposition of a hypothetical shape of steady-state rounding to a width *W* for a semi-infinite plate eroded under the same erosion conditions. When the material attempts to round to width *W,* a sharp cusp will result if $d < 2W$, but the cusp will be subject to easy fracture, with the result that no steady-state shape exists and *E* is increased. If *W* increases or *d* decreases, *E* increases further. For SPPs larger than $d_{SPP} = 2W$, the edge effect should vanish and the second phase should retain its high inherent erosion resistance. *W* increases with the size of chips generated by impacting particles and therefore with the kinetic energy of the particles, and as *W* increases, the magnitude of the edge effect increases.

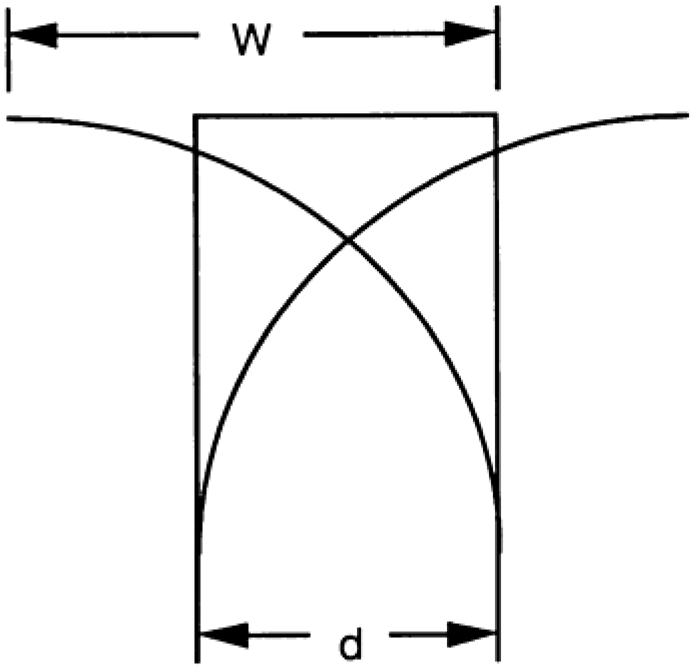


Fig. 10 Schematic showing square edge of a plate, *d,* or end of a rod of diameter d_r, being eroded from the top. Curves show superposition of a steady-state shape of rounding of width *W,* which would occur for the edge of a thick, semi-infinite plate.

Direct evidence for the edge effect was obtained by Ahmed (Ref 92) and Kosel (Ref 93), who eroded glass of a range of diameters with crushed quartz particles of varying size and velocity incident parallel to the rod axis. Flat plates of the same glass were also eroded at normal incidence under the same conditions; in the absence of the edge effect, the ratio of rod to plate erosion rate, E_r/E_p, would be unity. E_r, E_p, and *W* were measured, and it was found that E_r/E_p decreased from values as high as 70 to close to unity as rod diameter increased to about 2*W.* The E_r/E_p results for four velocities and particle sizes also fell on a single curve with a correlation coefficient of 0.93 when plotted against the ratio d_r/W, which characterizes the geometry in Fig. 10.

Related studies were also reviewed by Kosel and Ahmed (Ref 86). According to Ninham and Levy (Ref 94), smaller SPPs are often rendered ineffective by through-thickness fracture rather than the edge effect, and the subsurface SPP fracture is observed (Ref 94, 95, 96) and in some cases produces void nucleation. Kosel and Ahmed (Ref 86) argued that these phenomena are related to the edge effect by the lack of sufficient support offered by a relatively soft matrix.

A study by Rossing and Rocazella (Ref 97) of slurry erosion of alloys with a high volume fraction of large SiC SPPs in a matrix of Al/Al_2O_3 demonstrated that the edge effect is absent for sufficiently mild erosion conditions (Ref 86, 97), so that such materials can approach the erosion resistance of a monolithic ceramic. This is a particularly important result, because it demonstrates that there are erosion condition under which the edge effect can be avoided. The conditions under which the edge effect disappears are predicated by the edge-effect concepts given above. The key is that low-velocity erosion, such as occurs in slurry erosion, generates such small chip dimension at SPP edges that with sufficiently large SPPs, the condition $d_{SPP} > 2W$ can be obtained. Relatively ductile alloys with dispersed ceramic phases can then approach the much higher inherent erosion resistance of ceramics, provided the distance between SPPs is sufficient to prevent their undermining and direct removal.

The edge is also a significant factor in abrasion (Ref 98). However, it is of lower magnitude than in erosion, probably because abrasive particles approach the SPPs tangent to the surface.

WC-Co cermets represent an important class of materials in which the edge effect is not applicable. Erosion of such materials has been studied extensively (Ref 94, 99, 100) because of their generally very high erosion resistance relative to other alloys. These materials typically have very high SPVFs and very small SPPs, and the lack of SPP support by the adjacent matrix is complicated by several factors. The matrix yield strength is significantly higher due to the constraint of adjacent carbides and very small mean free path in the matrix. The carbides are so close that they often touch, creating a carbide "skeleton" that can transmit forces to adjacent carbides, and the carbides in WC-Co cermets are typically so small (1 to 3 μm) that an impacting particle will often contact more than one carbide. The combination of support from the constrained matrix and the carbide network, together with the very high SPVF, is probably responsible for the high erosion resistance of such materials.

Ninham and Levy (Ref 94) found a clear decrease of E with increasing SPVF for quartz erosion of WC-Co, WC-Ni, and related cermets. Anand and Conrad (Ref 100, 101) studied Al_2O_3 erosion of WC-Co and found a transition of the peak erosion angle from 90° to low angles as the scaling of the impact damage zone increased in size from one comparable to the WC grain size to one much larger. This was explained as a transition from brittle to ductile behavior, with brittle fracture of WC in the former case and deformation of the matrix controlling the latter. The erosion rate in the ductile regime was found to be proportional to the square root of the matrix mean free path; for constant WC grain size, this would imply a decrease of E with increasing SPVF. The lateral crack dimensions in pure WC (grain size, 1000 to 3000 μm) was significantly greater than the dimensions of the impact damage site of even the least erosion-resistant cermet under the same impact conditions (Ref 100), suggesting that crack length is limited by the ductile matrix, with impact energy being absorbed more by plastic deformation of cobalt than by fracture of WC.

Erosion / Corrosion

Erosion/corrosion (E/C) refers to the simultaneous, synergistic interactions between SPE and corrosion. In this general sense, discussion should logically include E/C synergism in slurry, cavitation, and liquid erosion, as well as gas-borne particle erosion. The present discussion is limited to the latter area, in which thick scales (oxides, sulfides, and so forth) can develop on metallic substrates, unlike slurry, cavitation, and liquid erosion. The discussion is further limited to E/C of metals.

Much progress has been made on the subject of gas-borne particle E/C of metals within the past several years, and a reasonably complete and generally accepted picture of the important features of the problem is emerging. Many of the details remain under intensive investigation, and the reader is encouraged to consult the current literature to keep abreast of rapid developments in the field. A number of papers on E/C were presented at a recent conference (Ref 102). Although quantitative prediction of E/C rates is impractical, progress has been made toward predicting and understanding divisions between E/C regimes.

Erosion/corrosion is of considerable technical importance in several types of applications, including coal gasification or liquefaction, steam turbines, jet turbines, and in the in-bed evaporator tubes, waterwalls, and convention pass surfaces of FBC systems.

An extensive program on E/C in the high-temperature, high-pressure conditions is coal gasification systems was surveyed by Sorrel (Ref 103). Affected components include cyclones, piping, valves, and heat exchangers. It was found that chromium was the most beneficial alloying element and the E/C rates were typically about two orders of magnitude higher than simple corrosion. Iron, nickel, and their oxides are thermodynamically unstable in the sulfidizing atmosphere, resulting in the formation of nonprotective sulfides that tend to become molten and accelerate attack at the necessary operating temperatures. This situation is the antithesis of the protective oxide scales that appear to offer some promise for other, less sulfidizing applications. The E/C problem in coal gasification pressure vessels is treated by lining the vessels with refractory brick (Ref 104). Much recent E/C research has investigated oxidizing atmospheres in which sulfides are not dominant in the scales, such as coal-fired boilers and gas turbines.

Types of Erosion/Corrosion

In the simplest terms, E/C can be separated into two regimes--dominated by erosion or corrosion, depending on which process occurs at the greater rate under the conditions of interest. For very low corrosion rates and/or highly erosive conditions, erosion dominates and basically pure erosion is expected. For high corrosion rates or mild erosion conditions, corrosion dominates. Between these extremes, important synergistic interactions can increase the metal thickness loss rate significantly beyond that due to either erosion or corrosion alone. It is important to point out that a number of studies (Ref 105, 106, 107, 108, 109, 110, 111) have shown the existence of a maximum E/C rate at an intermediate temperature within the synergistic regime between pure erosion and pure corrosion, while others have shown a rapid rise followed by a leveling, which may indicate that a maximum would exist at a temperature beyond the range investigated. Some examples are given in Fig. 11.

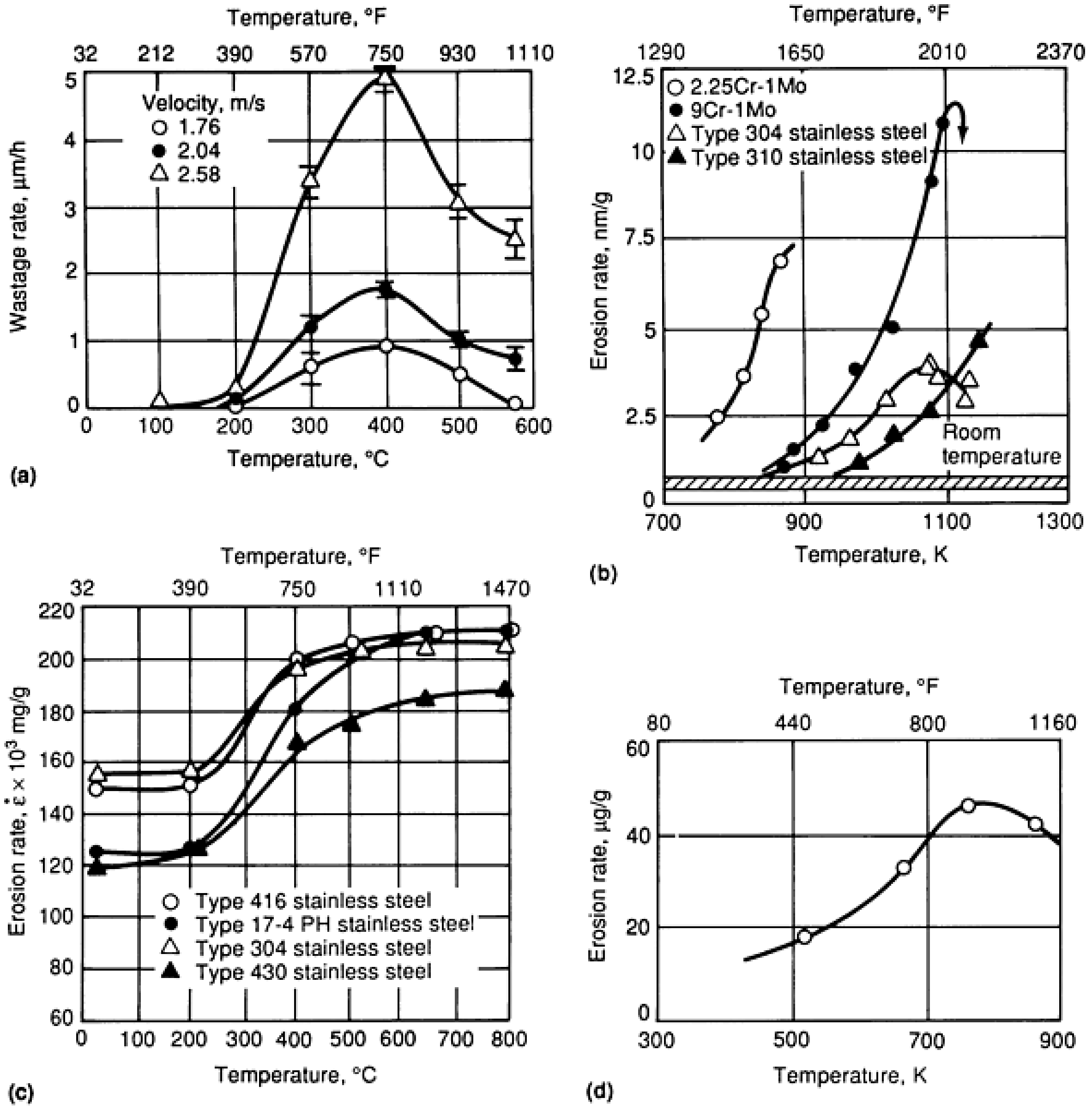


Fig. 11 Typical observed temperatures for maximum E/C rates of selected materials. (a) Fluidized bed combustion E/C of untempered martensitic type 401 stainless steel with 80-mesh alumina as a function of abrasive velocity. (b) Erosion/corrosion of four selected alloys by 1 mm Al_2O_3 particles at 30° angle of incidence and 2.7 m/s (8.9 ft/s) velocity. (c) Erosion/corrosion of selected stainless steels by 120-grit SiC at 30° angle of incidence and 65 m/s (215 ft/s) velocity. (d) Erosion/corrosion of 9Cr-1Mo at an incident angle of 30° and a velocity of 30 m/s (100 ft/s) with 250 μm SiC. Source: Ref 105, 107, 108, 129.

Most authors separate the synergistic regime between pure erosion and pure corrosion into two broad categories: (1) erosion-enhanced corrosion (EEC), in which the dimension of the impact damage zone is confined within a thick scale, and (2) corrosion-affected erosion (CAE), in which the damage zone includes both scale and metal. Erosion-enhanced corrosion causes increased metal thickness loss rates because erosion reduces the scale thickness and therefore increases the corrosion rate, which is controlled by transport of metal or corrosive elements (oxygen, sulfur, and so on). Scale spallation can also be involved. Corrosion-affected erosion often produces a surface composite consisting of intimately intermixed metal, scale, and embedded erodent particle fragments. Although it appears to produce the highest loss rates, the mechanisms of CAE have not yet been fully elucidated. A more detailed discussion of mechanisms and the effects of important variables follows. Note that the concern is always with the *relative* rates of erosion and corrosion, although the erosion component may refer either to erosion of the scale or of the metal substrate. The most logical sequence of increasing severity of corrosion relative to erosion is as follows: (1) erosion, (2) CAE, (3) EEC, and (4) corrosion. An important fact for applications involving E/C is that as temperature increases, the corrosion rate increases, resulting in

progression through the series just described. There is considerable evidence to suggest that the maximum E/C rate (Fig. 11) occurs at the transition between CAE and EEC, with the loss rate decreasing after this because of protection of the metal by the scale.

Rishel *et al.* (Ref 112) extended a previous E/C classification by Kang *et al.* (Ref 113, 114) to include scale spalling. They described the following four categories of E/C, which have been rearranged here in order of increasing corrosion rate. Figure 12 schematically illustrates the different regimes.

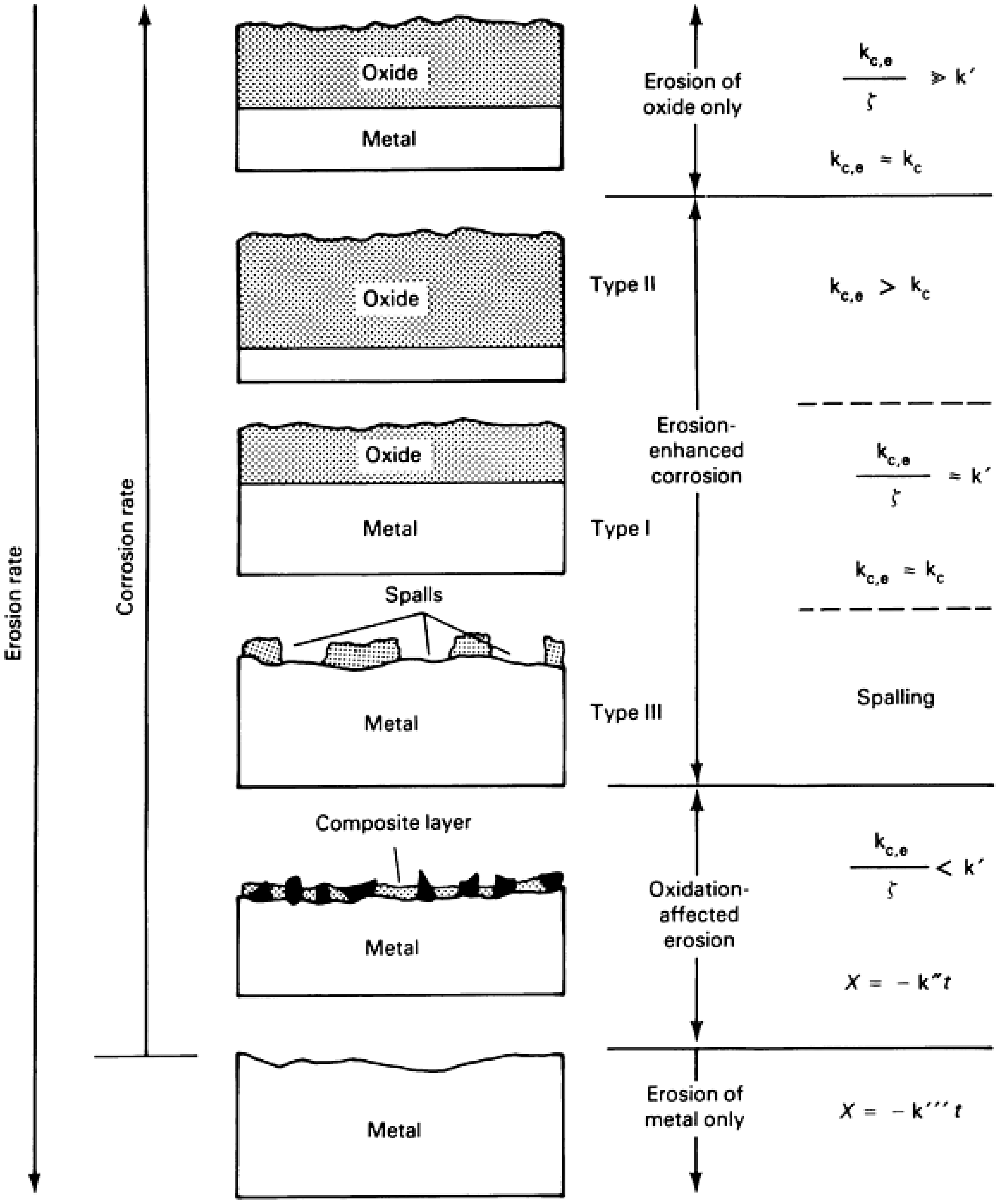


Fig. 12 Erosion/corrosion interaction regimes shown in order of increasing severity of corrosion relative to erosion. Types I, II and III regimes are subdivisions of EEC.

Pure erosion occurs when the corrosion rate is negligible compared with the erosion rate, as in severe erosion conditions (high-velocity, angular particles) or in noncorroding conditions such as an inert gas. In a corrosive gas, this regime would be confined to relatively low temperatures. The mechanisms and phenomenology of erosion of metals discussed earlier apply.

Corrosion-Affected Erosion. With an increased corrosion component, a thin scale is developed, but the dimensions of the stress field introduced by the particles are greater than the scale thickness, so that both scale and metal are deformed. The metal loss rate is increased over that for pure erosion, and a surface composite of metal, scale, and erodent particle fragments is typically produced. The mechanisms by which the E/C synergism increases the metal los rate in this regime appear to be the least fully understood, as discussed later. One likely explanation is that bare metal is repeatedly exposed to corrosion. A parabolic rate equation would predict an infinite initial corrosion rate, but is only applicable after the early stage of scale growth, which is controlled by other mechanisms (Ref 115).

Erosion-enhanced corrosion was broken into three categories. The first two categories involve a steady-state scale thickness, ζ_{ss}, determined by the relative erosion and corrosion rates, and the third involves spalling of the scale.

Metals likely to be of use in high-temperature applications in air or other corrosive atmospheres are those that form protective scales that grow naturally in a state of compression, because the volume of scale formed from a given amount of metal exceeds that of the metal consumed (Ref 115). The compressive stress combats fracture, resulting in a protective scale free of cracks and pores. Scales that form in tension generally crack and are nonprotective. After a short initial stage, protective scale growth is limited by the diffusion of either metal outward or of the oxidant (oxygen, sulfur, halogen) inward through the scale thickness. Because the diffusion gradient is inversely proportional to ζ, scale thickness ζ is often proportional to $t^{0.5}$, referred to as parabolic oxidation. The protective scale growth rate, $d\zeta/dt$, during free corrosion is given by:

$$\frac{d\zeta}{dt} = \frac{k_p}{\zeta} \qquad \textbf{(Eq 4)}$$

where k_p is the parabolic rate constant.

During EEC, if a continuous protective scale is maintained without spalling, scale is removed from the outer surface at a rate k' (which will depend on erosion conditions). The net scale growth rate will then be (Ref 112, 113, 114):

$$\frac{d\zeta}{dt} = \frac{k_{c,e}}{\zeta} - k' \qquad \textbf{(Eq 5)}$$

where k_p has been replaced by $k_{c,e}$, the applicable parabolic rate constant during E/C. A steady-state scale thickness, ζ_{ss}, is possible when $d\zeta/dt = 0$, which gives:

$$\zeta_{ss} = \frac{k_{c,e}}{k'} \qquad \textbf{(Eq 6)}$$

If a specimen is exposed to simultaneous E/C, scale builds up to thickness ζ_{ss}, while if it is preoxidized to a thickness $\zeta > \zeta_{ss}$ and then erosion begins, $d\zeta/dt$ is initially negative and ζ decreases until ζ_{ss} is reached (Ref 112).

Type I EEC was defined by Rishel *et al.* (Ref 112) as the case just described, with $k_{c,e} = k_p$, that is with the scaling rate essentially unaffected by the presence of erosion. Good agreement was found (Ref 112, 116) between literature values of k_p and those calculated from measured values of ζ_{ss} and k', for E/C of Co.

Type II EEC was defined (Ref 112) by $k_{c,e} > k_p$, that is, the case in which the presence of erosion actually increases the scaling rate. This was observed by Rishel *et al.* (Ref 117) for commercial-purity nickel (Ni 270) in air + SO_2/SO_3, in which the scale thickness after E/C was about 50% greater than for free corrosion, and by Levy *et al.* (Ref 118) for low-chromium steels in methane/air mixtures. In the case of Ni 270, the effect $k_{c,e} > k_p$ was attributed to a redistribution of the

nickel sulfide network, which aids transport through the mixed oxide/sulfide scale. In general, $k_{c,e} > k_p$ could occur through mechanisms by which erosion alters the type or distribution of phases within the scale or introduces fast transport paths, such as dislocations or cracks.

Type III EEC was described (Ref 112) as the case in which erosion-induced spalling of the scale enters as an additional scale loss mechanism, resulting in a very thin or absent scale and a high corrosion rate. Spalling was included within the EEC regime (Ref 112) because it is a mechanism of discontinuous scale thinning that leads to an enhanced corrosion rate. The metal loss rate can be very sensitive to particle loading rate (in $g/cm^2 \cdot s$); a threefold increase of loading rate produced a 13-fold increase of type III EEC rate for nickel. (In the absence of spallation, the metal loss rate was simply proportional to loading rate.) In the absence of erosion, protective scales tend to spall above a certain thickness due to buckling induced by the natural compressive stresses, but this requires separation at the scale/metal interface.

Rishel *et al.* (Ref 112) discussed work by van der Zwaag and Field (Ref 110) that showed that indentation on a coating can significantly increase the maximum shear stress at the coating/substrate interface, promoting separation that could lead to spalling by buckling. Levy (Ref 120) suggested that spalling of scale in E/C of 9Cr-1Mo steel could be attributed to consolidation of the scale due to hot pressing action caused by particle impact. This was believed to eliminate nodule boundaries, forming a continuous scale susceptible to buildup of stresses and thus spalling. Other work by Rishel *et al.* (Ref 117) showed spalls on Ni 270 of sizes comparable to the impact dimension; this may be due to stress wave effects (Ref 112).

Pure Corrosion. Rishel *et al.* (Ref 112) describe this regime as one in which the ratio of corrosion to (scale) erosion rate is very high, resulting in pure corrosion and parabolic scale thickening. However, for a thick scale in the presence of a flux of erodent particles, it is important to realize that unless the scale erosion rate k', is actually zero, there must always be a steady-state scale thickness ζ_{ss} as described earlier. Even though this thickness may be large, a *constant* rate of metal thickness loss would be expected with a constant scale thickness, whereas for free corrosion (k' = 0) the rate of metal thickness loss decreases continuously unless scale buckling and spalling occur above a critical thickness. Certain, if the metal loss rate is sufficiently low, the distinction is of little practical significance.

Time- Versus Mass-Based E/C Rates

Sethi and Wright (Ref 111, 121) discussed the importance of flux in E/C, and based on a simple model in which the entire scale thickness is removed down to bare metal within the "footprint" of each impact, they demonstrated the different behavior expected when *E* is expressed in units of mass loss per unit time (time-based) as opposed to mass loss per unit mass impacted (mass-based). Because the erosion rate varies with flux while the corrosion rate does not (except for synergistic interactions, such as effects of erosion on scale properties), flux should be important through its effect on relative erosion and corrosion rates, regardless of the model assumed for E/C synergism.

Figure 13 illustrates how a steady-state E/C rate would occur based on the model assumed for simplicity by Sethi and Wright (Ref 121), for fluxes $F_1 = 4F_3 = 16F_5$. For higher fluxes, less time elapses between scale-removal events, and even though the amount (thickness) of scale removed per impact is less, the overall time-based E/C rate increases as $F^{0.5}$ due to the higher average parabolic corrosion rate at short times. In contrast, the mass-based E/C rate would be proportional to $F^{-0.5}$, as was observed by Sethi and Corey (Ref 108) for 2.25Cr-1Mo steel. Raising the temperature increases the E/C rate because of the increase of corrosion rate. The authors observed that this mechanism is not always applicable, but that time- and mass-based units will also give different flux effects for other mechanisms. The authors noted that if the oxide thickness grown in the fixed time between impacts exceeds a critical value at a higher temperature, the system will enter the scale erosion regime, and the E/C rate will be expected to decrease. They noted several studies in which a maximum E/C rate has been observed at some temperature.

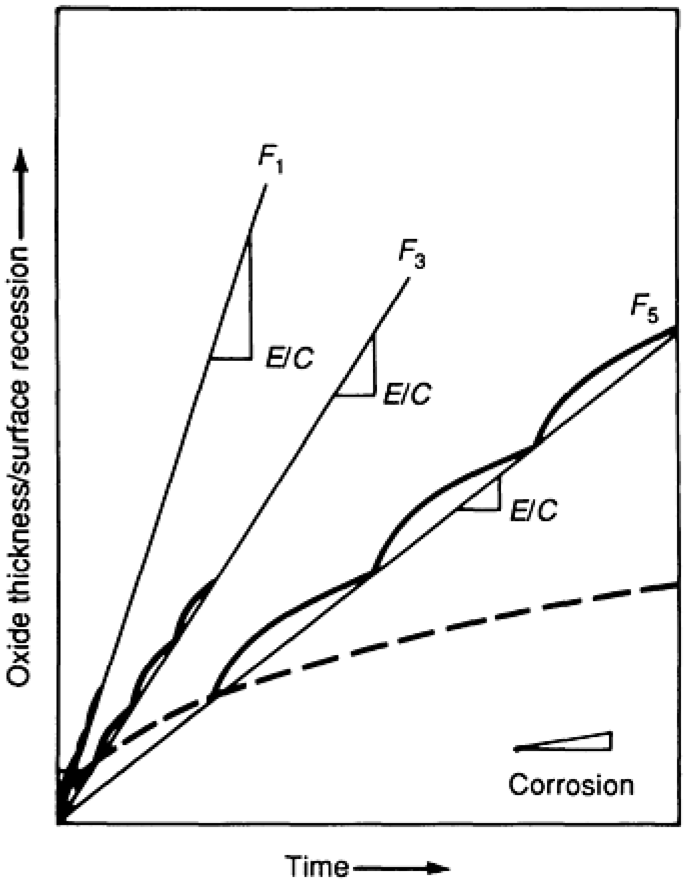


Fig. 13 Plot of oxide thickness versus time as a function of flux, *F*, at temperature T_3. The assumed parabolic corrosion curve is also shown. Source: Ref 111

Comparison With Other E/C Classifications

An earlier study by Hogmark *et al.* (Ref 122) discussed a classification scheme similar to that of Rishel *et al.* (Ref 112) as originated by Kang *et al.* (Ref 114). In order of increasing relative corrosion rate, they listed (1) pure erosion, (2) simultaneous damage of metal and scale (equivalent to CAE), (3) flaking (spalling) of scale at individual impacts, (4) erosion of scale without flaking, (5) erosion-affected corrosion, and (6) pure corrosion. Pure corrosion was treated as involving periodic flaking due to stress buildup, and erosion-affected corrosion was attributed to effects of erosion on the critical thickness for flaking or spalling. The regime of erosion of scale without flaking was not treated in terms of a steady-state scale thickness as in the work of Rishel *et al.* (Ref 112).

Sundararajan (Ref 49, 123) has classified E/C in a manner similar in most respect to that of Rishel *et al.* (Ref 122). One difference is that he separates what Rishel *et al.* termed type III EEC (scale spalling) into two separate cases of "oxidation-controlled" erosion: (1) continuous oxide removal, in which all oxide is removed down to bare metal within each impact site, and (2) intermittent spalling of oxide when the scale builds up to a critical thickness, ζ_c, only after several impacts. Sundararajan's treatment of the effects of the several important variables in the various regimes and his application of wear maps to their interpretation is quite useful. He also discussed the role of increased plasticity of scale along with the effect of increasing scale thickness with temperature. A ductile-to-brittle transition (DBT) occurs in many protective oxide scales; Stephenson *et al.* (Ref 124) observe that single impacts on oxide scales on nickel-base superalloys gave brittle behavior below a critical temperature for a fixed scale thickness and that, at high temperature, impacts went from ductile to brittle behavior with increasing scale thickness. Thus, Sundararajan (Ref 49) defined a critical scale thickness above scale behavior became brittle. Hutchings *et al.* (Ref 105) also attributed the decreasing E/C rate above the maximum temperature to the combined effects of increasing scale growth rate and scale plasticity with increasing temperature.

Erosion-Enhanced Corrosion Regime

If the picture presented by Rishel *et al.* (Ref 112) of a steady-state scale thickness being achieved by equal scale growth and loss rates in types I and II EEC is valid, as it appears to be from the published results, a simple point should be made with regard to the implications for the temperature dependence of the metal thickness loss rate. Because it is postulated that $k_{c,e}/\zeta = k'$, the scale growth rate $k_{c,e}/\zeta$ is directly controlled by the scale erosion rate k', because erosion controls ζ_{ss} at a given temperature and therefore controls the rate of diffusion through the scale. The metal thickness loss rate is proportional to the scale growth rate through appropriate density factors. This being true, even though $k_{c,e}$ varies significantly with temperature, the condition that growth and loss rates are equal implies that the temperature dependence of the metal thickness loss is controlled by k', not $k_{c,e}$. This point does not seem to have been made by other authors. It also follows that if the scale erosion rate were independent of temperature in the EEC temperature regime, the metal thickness loss rate would remain constant. As described above, other authors have alluded to the effect of increasing scale ductility with temperature as decreasing its erosion rate, and it appears that this may be responsible for the fact that the E/C rate is observed to reach a maximum and then decrease with temperature.

Scale Flaking or Spalling

Scale spalling is included by most authors as an important mechanism, but there is lack of general agreement as to whether it belongs in the CAE or EEC regime, or where it falls with respect to the often observed maximum E/C temperature. Sethi and Wright (Ref 111, 121) discussed intermittent parabolic growth between spalling events, which they considered a mechanism of CAE associated with rising E/C rate with increasing temperature. Rishel *et al.* (Ref 112) included it as type III EEC and presented an extensive discussion of spalling mechanisms. Sundararajan (Ref 49) included spalling as the oxidation-controlled erosion regime.

Levy *et al.* (Ref 118, 125, 126, 127, 128) showed the importance of erosion-induced changes of scale morphology on spalling. Segmented scales were found to give lower E/C rates compared with consolidated scales. Segmented scales formed on 9Cr-1Mo steel by growth of nodules that impinged (Ref 120); nodule growth is rare in free oxidation and was attributed to initiation of nodules by damage to thin scale at impact sites. Scale consolidation occurred at higher velocity and was attributed to compressive stresses generated by the normal component of velocity, which leads to a process similar to hot isostatic pressing at high enough E/C temperatures. The E/C rate increased suddenly by a factor of eight at 30 m/s (100 ft/s) when consolidated scale was formed, apparently because the area that spalls is limited to the original crystallite size in segmented but not in consolidated scales. The consolidation process is found to be more effective for round Al_2O_3 particles than for angular SiC particles. Also, higher chromium content promotes the segmented scale morphology (Ref 125, 128) and therefore gives lower E/C rates.

Corrosion-Affected Erosion Regime

There is lack of general agreement on the mechanisms of CAE, and the process appears difficult to model. Sethi and Wright's model of periodic scale-removal events with intervening periods of scale growth (Ref 111, 121) differs significantly from the description of Rishel *et al.* (Ref 112) of a continuous removal of a composite layer consisting of deformed metal, corrosion products, and embedded erodent particle fragments. In the model of Rishel *et al.*, the removal rate of the composite is the controlling factor, while in Sethi and Wright's model, the E/C rate is controlled entirely by the rapid mean corrosion rate for the thin scale that builds up between impact events. Rishel *et al.* also discuss the importance of including transient oxidation characteristics, since the scale is thin or absent. For Sethi and Wright's simplified assumptions, the rate of CAE should be close to the mean oxidation rate for the time between impacts, but this has not been established. As Sethi and Wright point out, some metal may be removed along with oxidation products, as would be expected when the depth of particle penetration exceeds the thickness of scale built up between impacts. It has been established that the majority of material removal in erosion of pure metals involves multiple impact phenomena, and it is likely that this would also be true in the CAE regime, unless scale adherence is weak and spalling is easy in a particular case. A composite layer such as that discussed by Rishel *et al.* might thus be expected. Because there are no satisfactory quantitative models for the aspects of material removal involving multiple impacts in pure erosion, quantitative modeling of the removal rate of the composite metal/scale layer appears very difficult.

Sundararajan (Ref 49) also refers to composite layer formation, and in addition notes that if the scale is ductile enough, it can partially shield the metal. Ductile response of scales to impact has been observed by Zhou and Bahadur (Ref 106, 129) and by Stephenson *et al.* (Ref 124). Oxide plasticity is expected at higher temperatures, and more brittle behavior may be expected in low-temperature CAE.

Despite the difficulty of quantitatively modeling CAE, it is worth discussing some of the factors that could be involved. Enhancements of the corrosion rate could occur by pipe diffusion along dislocations introduced by plastic deformation of

the metal or scale, or by short-circuit diffusion along metal/scale fragment and metal/embedded erodent fragment interfaces produced within a composite layer. One way in which corrosion could significantly increase the metal erosion rate is by penetration of corrosion through the highly deformed, thin platelets or displaced lips of metal, so that they are more easily detached. This would be particularly effective at lower temperatures, where corrosion products are brittle. However, this mechanism depends on whether displaced metal lips are formed, which may occur infrequently in a composite layer. Also, a thin scale overlying the metal might prevent rewelding of extruded platelets to the surface, increasing the metal erosion rate due to the presence of oxidation.

Sundararajan (Ref 49) estimates the range of conditions over which CAE should dominate. He assumes that if the scale thickness is greater than one-tenth of the depth of penetration in the metal, the scale will begin to influence the process promoting CAE, while if the scale thickness is greater than twice the depth of penetration in the scale, EEC will prevail. This leads to the conclusion that there should be only a very narrow range of velocity in which ζ_{ss} is in the correct range to produce CAE. His estimates also give a narrow range of temperature (about 100 °C, or 210 °F) over which a particular velocity would produce CAE. Experiments are in general agreement with this, often showing strong increases in E/C rates over a range of 100 to 200 °C (210 to 390 °F) (Ref 107, 108, 110, 111) before a maximum is reached, although in some cases the temperature range between pure erosion and the maximum is closer to 300 °C (570 °F) (Ref 105, 106, 109).

References

1. D. J. Buquoi, "Solid Particle Erosion: Its Causes, Effects and Solutions," presented at Florida Municipal Electric Association 1989 Engineering & Operations Workshop, Altamonte Springs, 31 Oct-2 Nov 1989
2. T.H. McCloskey and C. Bellanca, Minimizing Solid Particle Erosion in Power Plant Steam Turbines, *Power Eng.,* Aug 1981, p 35-38
3. W. Tabakoff, Effect of Environmental Particles on a Radial Compressor, Paper 26, *Proceedings of Corrosion-Erosion-Wear of Materials at Elevated Temperatures,* A. V. Levy, Ed., NACE/EPR/LBL/DOE-FE, Berkeley, CA, 31 Jan-2 Feb 1990
4. I. Finnie, The Mechanisms of Erosive Wear in Ductile Metals, *Corrosion-Erosion Behavior of Materials,* K. Natesan, Ed., TMS-AIME, 1980, p 118-126
5. J.S. Hansen, Relative Erosion Resistance of Several Metals, *Erosion: Prevention and Useful Applications,* STP 664, ASTM, 1979, p 148-162
6. "Standard Practice for Conducting Erosion Tests by Solid Particle Impingement Using Gas Jets," G76-83, ASTM
7. I. Finnie, The Mechanism of Erosion of Ductile Metals, *Proceeding of 3rd U.S. National Congress of Applied Mechanics,* 1958, p 527-532
8. I. Finnie, Erosion of Surfaces by Solid Particles, *Wear,* Vol 3, 1960, p 87-103
9. I. Finnie and D.J. McFadden, On the Velocity Dependence of the Erosion of Ductile Metals by Solid Particles of Low Angles of Incidence, *Wear,* Vol 48, 1978, p 181-190
10. I. Finnie, J. Wolak, and Y. Kabil, Erosion of Metals by Solid Particles, *J. Mater.,* Vol 2, 1967, p 682-700
11. K.C. Goretta, R.C. Arroyo, C.-T. Wu, and J.L. Routbort, Erosion of Work-Hardened Copper, Nickel and 304 Stainless Steel, *Wear,* Vol 147, 191, p 145-154
12. I.M. Hutchings, R.E. Winter, and J.E. Field, Solid Particles Erosion of Metals: The Removal of Surface Material by Spherical Projectiles, *Proc. R. Soc. (London) A,* Vol 348, 1976, p 379-392
13. I.M. Hutchings, Deformation of Metal Surfaces by the Oblique Impact of Square Plates, *Int. J. Mech. Sci.,* Vol 19, 1977, p 45-52
14. R.E. Winter and I.M. Hutchings, Solid Particles Erosion Studies Using Single Angular Particles, *Wear,* Vol 29, 1974, p 181-194
15. I.M. Hutchings, Mechanisms of the Erosion of Metals by Solid Particles, *Erosion: Prevention and Useful Applications,* STP 664, ASTM, 1979, p 59-76
16. I. Finnie, A.V. Levy, and D.J. McFadden, Fundamental Mechanisms of the Erosive Wear of Ductile Metals by Solid Particles, *Erosion: Prevention and Useful Applications,* STP 664, ASTM, 1979, p 36-58
17. I. Finnie, Erosion of Metals, *Proceedings of Conference on Corrosion/Erosion of Coal Conventional System Materials,* National Association of Corrosion Engineers, 1979, p 429-443

18. R.A. Mayville, "Mechanism of Material Removal in the Solid Particle Erosion of Ductile Metals," M.S. thesis, University of California--Berkeley, 1978
19. R. Bellman, Jr. and A. Levy, Platelet Mechanism of Erosion of Ductile Metals, *Proceedings of International Conference on Wear of Materials,* American Society of Mechanical Engineers, 1981, p 564-576
20. A.V. Levy, The Erosion of Metal Alloys and Their Scales, *Proceedings of Conference on Corrosion-Erosion-Wear of Materials in Emerging Fossil Energy Systems,* National Association of Corrosion Engineers, 1982, p 298-376
21. T.H. Kosel, Z.Y. Mao, and S.V. Prasad, Erosion Debris Particle Observations and the Micromachining Mechanism of Erosion, *ASLE Trans.,* Vol 28, 1984, p 268-276
22. A. Hammarsten, S. Soderberg, and S. Hogmark, Study of Erosion Mechanisms of Recovery and Analysis of Wear Fragments, *Proceedings of International Conference on Wear of Materials,* American Society of Mechanical Engineers, 1983, p 373-381
23. A.V. Levy, The Platelet Mechanism of Erosion of Ductile Metals, *Wear,* Vol 108, 1986, p 1-21
24. I.M. Hutchings and A.V. Levy, Thermal Effects in the Erosion of Ductile Metals, *Proceedings of Conference on Corrosion-Erosion-Wear of Materials at Elevated Temperatures,* A.V. Levy, Ed., NACE/EPRI/LBL/DOE-FE, Berkeley, CA, 27-29 Jan 1986, p 121-133
25. J.G.A. Bitter, A Study of Erosion Phenomena, Part I, *Wear,* Vol 6, 1963, p 5-21
26. J.G.A. Bitter, A Study of Erosion Phenomena, Part II, *Wear,* Vol 6, 1963, p 169-190
27. J.H. Neilson and A. Gilchrist, Erosion by a Stream of Solid Particles, *Wear,* Vol 11, 1968, p 111-122
28. C.T. Morrison, R.O. Scattergood, and J.L. Routbort, Erosion of 304 Stainless Steel, *Wear,* Vol 111, 1986, p 1-13
29. D.G. Rickerby and N.H. Macmillan, Erosion of Aluminum and Magnesium Oxide by Spherical Particles, *Proceedings of International Conference on Wear of Materials,* American Society of Mechanical Engineers, 1981, p 548-563
30. D.G. Rickerby and N.H. Macmillan, On the Oblique Impact of a Rigid Sphere Against a Rigid-Plastic Solid, *Int. J. Mech. Sci.,* Vol 22, 1980, p 491-494
31. I.M. Hutchings, D.G. Rickerby, and N.H. Macmillan, Further Studies of the Oblique Impact of a Hard Sphere Against a Ductile Solid, *Int. J. Mech. Sci.,* Vol 23, 1981 p 639-646
32. R. Brown, E. Jin Jun, and J.W. Edington, Mechanisms of Erosive Wear for 90° Impact on Copper and Iron Targets, *Proceedings of International Conference on Wear of Materials,* American Society of Mechanical Engineers, 1981, p 583-591
33. J. Maji and G.L. Sheldon, Mechanisms of Erosion of a Ductile Material by Solid Particles, *Erosion: Prevention and Useful Applications,* STP 664, ASTM, 1979, p 136-147
34. S. Bahadur and R. Badruddin, Erodent Particle Characterization and the Effect of Particle Size and Shape on Erosion, *Proceedings of International Conference on Wear of Materials,* American Society of Mechanical Engineers, 1989, p 143-153
35. J.W. Edington and I.G. Wright, Study of Particle Erosion Damage in Haynes Stellite 6B, I: Scanning Electron Microscopy of Eroded Surfaces, *Wear,* Vol 48, 1978, p 131-144
36. J.W. Edington and I.G. Wright, Study of Particle Erosion Damage in Haynes Stellite 6B, II: Scanning Electron Microscopy of Eroded Surfaces, *Wear,* Vol 48, 1978, p 145-155
37. L.K. Ives and A.W. Ruff, Electron Microscopy Study of Erosion Damage in Copper, in *Erosion: Prevention and Useful Applications,* STP 664, ASTM, 1979, p 5-35
38. A.W. Ruff and S.M. Wiederhorn, Erosion by Solid Particle Impact, *Treat. Mater. Sci. Technol.,* Vol 16, 1979, p 69-126
39. T.H. Kosel, R.O. Scattergood, and A.P.L. Turner, An Electron Microscope Study of Erosive Wear, *Proceedings of International Conference on Wear of Materials,* American Society of Mechanical Engineers, 1979, p 192-204
40. T.H. Kosel, A.P.L. Turner, and R.O. Scattergood, Effects of Particle Size and Shape on Erosive Wear

Mechanisms, in *Corrosion-Erosion Behavior of Materials,* TMS-AIME, 1980, p 146-161

41. A. Misra ad I. Finnie, On the Size Effect in Abrasive and Erosive Wear, *Wear,* Vol 65, 1981, p 359-373
42. N. Gane and J.M. Cox, The Microhardness of Metals at Very Low Loads, *Philos. Mag.,* Vol 22, 1970, p 881-891
43. N. Gane, The Direct Measurement of the Strength of Metals on a Sub-Micrometer Scale, *Proc. R. Soc. (London) A,* Vol 317, 1970, p 367-391
44. M. Liebhard and A.V. Levy, The Effect of Erodent Particle Characteristics on the Erosion of Metals, *Proceedings of International Conference on Wear of Materials,* American Society of Mechanical Engineers, 1991, p 123-127
45. R.C.D. Richardson, The Wear of Metals by Relatively Soft Abrasives, *Wear,* Vol 11, 1968, p 245-275
46. R.C.D. Richardson, The Wear of Metals by Hard Abrasives, *Wear,* Vol 10, 1967, p 291-309
47. K. Wellinger and H. Uetz, Gleit-Spül and Strahlverschleiss Prüfung, *Wear,* Vol 1, 1957-58, p 225-231
48. K. Anand, S.K. Hovis, H. Conrad, and R.O. Scattergood, Flux Effects in Solid Particles Erosion, *Wear,* Vol 118, 1987, p 243-257
49. G. Sundararajan, The Solid Particle Erosion of Metallic Materials at Elevated Temperatures, Paper 11, *Proceedings of Conference on Corrosion-Erosion-Wear of Materials at Elevated Temperatures,* A.V. Levy, Ed., NACE/EPRI/LBL/DOE-FE, Berkeley, CA, 31 Jan-2 Feb 1990
50. I. Finnie, An Experimental Study of Erosion, *Proc. Soc. Exp. Stress Anal.,* Vol 17, 1960, p 65-70
51. G.A. Sargent, M.K. Keshavan, P.K. Mehrotra, and H. Conrad, The Erosion of Plain Carbon Steels by Ash Particles From a Coal Gasifier, *Proceedings of International Conference on Wear of Materials,* American Society of Mechanical Engineers, 1981, p 613-618
52. L.G. Peterson, G.A. Sargent, and H. Conrad, Effect of Microstructure on the Erosion of Steel by Solid Particles, *Proceedings of International Conference on Wear of Materials,* American Society of Mechanical Engineers, 1985, p 661-670
53. G.L. Sheldon and I. Finnie, On the Ductile Behavior of Nominally Brittle Materials During Erosive Cutting, *Trans. ASME,* Vol 88B, 1966, p 387-392
54. I. Finnie and Y.H. Kabil, On the Formation of Surface Ripples During Erosion, *Wear,* Vol 8, 1965, p 60-69
55. J. Salik and D.H. Buckley, Effects of Erodent Particle Shape and Various Heat Treatments on Erosion Resistance of Plain Carbon Steels, Technical Paper 1755, National Aeronautics and Space Administration, 1981
56. A.V. Levy and S. Jahanmir, The Effects of the Microstructure of Ductile Alloys on Solid Particle Erosion Behavior, in *Corrosion-Erosion Behavior of Materials,* TMS-AIME, 1980, p 177-189
57. W. Zhu and Z.Y. Mao, Study of Erosion by Relatively Soft Particles, *Proceedings of International Conference on Wear of Materials,* American Society of Mechanical Engineers, 1987, p 787-796
58. X. Liang and T.H. Kosel, unpublished work
59. A.W. Ruff and S.M. Wiederhorn, Erosion by Solid Particle Impact, *Treat. Mater. Sci. Technol.,* Vol 16, 1979, p 69-126
60. S.M. Wiederhorn, Erosion of Ceramics, *Proceedings of Conference on Corrosion/Erosion of Coal Conventional System Materials,* National Association of Corrosion Engineers, 1979, p 444-479
61. A.G. Evans, Impact Damage Mechanics: Solid Projectiles, *Treat. Mater. Sci. Technol.,* Vol 16, 1979, p 1-67
62. J. Routbort and R.O. Scattergood, Solid Particle Erosion of Ceramics and Ceramic Composites, *Erosion of Ceramic Materials,* J.E. Ritter, Ed., Trans Tech, 1992, in press
63. J.E. Ritter, Ed., *Erosion of Ceramic Materials,* Trans Tech, 1992, in press
64. B.R. Lawn and T.R. Wilshaw, *J. Matter. Sci.,* Vol 10, 1975, p 1049
65. P.K. Mehrotra, G.A. Sargent, and H. Conrad, A Model for the Multiparticle Erosion of Brittle Solids by Spherical Particles, *Corrosion-Erosion Behavior of Materials,* TMS-AIME, 1980, p 127-145

66. B.J. Hockey, S.M. Wiederhorn, and H. Johnson, Erosion of Brittle Materials by Solid Particle Impact, *Fracture Mechanics of Ceramics,* Vol 3, *Flaws and Testing,* Plenum Press, 1978, p 379-402
67. H.R. Shetty, T.H. Kosel, and N.F. Fiore, A Study of Abrasive Wear Mechanisms Using Diamond and Alumina Scratch Tests, *Wear,* Vol 80, 1982, p 347-376
68. A.G. Evans, M.E. Gulden, and M. Rosenblatt, Impact Damage in Brittle Materials in the Elastic-Plastic Response Regime, *Proc. R. Soc. (London) A,* Vol 361, 1978, p 343-365
69. S.M. Wiederhorn and B.R. Lawn, Strength Degradation of Glass Impacted with Sharp Particles: I, Annealed Surfaces, *J. Am. Ceram. Soc.,* Vol 62, 1979
70. J.L. Routbort, R.O. Scattergood, and E.W. Kay, The Erosion of Si Single Crystals, *J. Am. Ceram. Soc.,* Vol 63, 1980, p 635
71. S. Srinivasan and R.O. Scattergood, On Lateral Cracks in Glass, *J. Mater. Sci.,* Vol 22, 1987, p 3463
72. R.O. Scattergood and J.L. Routbort, Velocity Exponent in Solid-Particle Erosion of Si, *J. Am. Ceram. Soc.,* Vol 66, 1983, p C184
73. L.M. Murugesh and R.O. Scattergood, Effect of Erodent Properties on the Erosion of Alumina, *J. Mater. Sci.,* Vol 26, 1991, p 5456-5466
74. J.L. Routbort, D.A. Helberg, and K.C. Goretta, Erosion of Ceramic Matrix Composites, *J. Hard Mater.,* Vol 1, 1990, p 123
75. S. Wada and N. Watanabe, Solid Particle Erosion of Brittle Materials, Part 7: The Erosive Wear of Commercial Ceramic Tools, *J. Ceram. Soc. Jpn. Int. Ed.,* Vol 96, 1988, p 323
76. J.E. Ritter, K. Jakus, M. Viens, and K. Breder, Effect of Microstructure on Impact Damage of Polycrystalline Alumina, Paper 55, *Proceedings of 7th International Conference on Erosion by Liquid and Solid Impact,* Cambridge University Press, 1987
77. J.E. Ritter, L. Rosenfeld, and K. Jakus, Erosion and Strength Degradation in Alumina, *Wear,* Vol 111, 1986, p 335-346
78. J.E. Ritter, Erosion Damage in Structural Ceramics, *Mater. Sci. Eng.,* Vol 71, 1985, p 195
79. J.E. Ritter, P. Strzepa, K. Jakus, L. Rosenfield, and K.J. Buckman, Erosion Damage in Glass and Alumina, *J. Am. Ceram. Soc.,* Vol 67, 1984, p 769
80. K. Breder, J.E. Ritter, and K. Jakus, Strength Degradation in Polycrystalline Alumina Due to Sharp-Particle Impact Damage, *J. Am. Ceram. Soc.,* Vol 71, 1988, p 1154
81. S. Srinivasan, "Erosion of Partially Stabilized Zirconia," M.S. thesis, North Carolina State University, 1987
82. M.E. Gulden, Solid-Particle Erosion of High-Technology Ceramics (Si_3N_4, Glass-Bonded Al_2O_3, and MgF_2), in *Erosion: Prevention and Useful Applications,* STP 664, ASTM, 1979, p 101-122
83. M.E. Gulden, Solid Particle Erosion of Si_3N_4 Materials, *Wear,* Vol 69, 1981, p 115-129
84. J.L. Routbort, C.-Y. Chu, J.M. Roberts, J.P. Singh, W. Wu, and K.C. Goretta, Erosion of Ceramic Composites by Various Erodents, Paper 31, *Proceedings of Conference on Corrosion-Erosion-Wear of Materials at Elevated Temperatures,* A.V. Levy, Ed., NACE/EPRI/LBL/DOE-FE, Berkeley, CA, 31 Jan-2 Feb 1990
85. A. Misra and I. Finnie, Correlations Between Two-Body and Three-Body Abrasion and Erosion of Metals, *Wear,* Vol 68, 1981, p 33-39
86. T.H. Kosel and T. Ahmed, The Edge Effect in Solid Particle Erosion of Ceramic Second-Phase Particles, *Erosion of Ceramic Materials,* J.E. Ritter, Ed., Trans Tech, 1992, in press
87. S.S. Aptekar and T.H. Kosel, Erosion of White Cast Irons and Stellite, *Proceedings of International Conference on Wear of Materials,* American Society of Mechanical Engineers, 1985, p 677-686
88. S.V. Prasad and T.H. Kosel, A Study of Carbide Removal Mechanisms During Quartz Abrasion, I: In-Situ Scratch Test Studies, *Wear,* Vol 92, 1983, p 253-268
89. S.V. Prasad and T.H. Kosel, A Comparison of Carbide Fracture During Fixed Depth and Fixed Load Scratch Tests, *Proceedings of International Conference on Wear of Materials,* American Society of Mechanical Engineers, 1985, p 59-66

90. T. Kulik, T.H. Kosel, and Y. Xu, Effect of Depth of Cut on Second-Phase Particle Fracture in Abrasion of Two-Phase Alloys, *Proceedings of International Conference on Wear of Materials,* American Society of Mechanical Engineers, 1989, p 23-34

91. T.H. Kosel and S.S. Aptekar, Effect of Hard Second-Phase Particles on the Erosion Resistance of Model Alloys, Paper 113, *Corrosion '86,* National Association of Corrosion Engineers, 1986

92. T. Ahmed, "Enhanced Removal at Edges or Brittle Materials During Solid Particle Erosion," M.S. thesis, University of Notre Dame, 1987

93. T.H. Kosel, "Erosion in Dual-Phase Microstructures," Final Report to U.S. Dept. of Energy, ORNL/Sub/83-43336C/01, Dec 1987

94. A.J. Ninham and A.V. Levy, The Erosion of Carbide-Metal Composites, *Proceedings of International Conference on Wear of Materials,* American Society of Mechanical Engineers, 1987, p 825-831

95. R. Brown and J.D. Ayers, Solid Particle Erosion of Al 6061 with a Laser Melted and TiC Particle Injected Surface Layer, *Proceedings of International Conference on Wear of Materials,* American Society of Mechanical Engineers, 1983, p 325-332

96. S.S. Aptekar and T.H. Kosel, Erosion of White Cast Irons and Stellite, *Proceedings of International Conference on Wear of Materials,* American Society of Mechanical Engineers, 1985, p 677-686

97. B.R. Rossing and M.A. Rocazella, Slurry Erosion of Silicon Carbide Particulate Reinforced Alumina Composites, *Proceedings of Corrosion-Erosion-Wear of Materials at Elevated Temperatures,* National Association of Corrosion Engineers, 1990, p 39-1 to 39-21

98. T. Kulik and T.H. Kosel, Effects of Second-Phase Particle Size and Edge Microfracture on Abrasion of Model Alloys, *Proceedings of International Conference on Wear of Materials,* American Society of Mechanical Engineers, 1989, p 71-82

99. K. Anand, C. Morrison, R.O. Scattergood, H. Conrad, J.L. Routbort, and R. Warren, "Erosion of Multiphase Materials," presented at 2nd International Conference on Science of Hard Materials (Rhodes, Greece), 1985

100. K. Anand and H. Conrad, Microstructure Effects in the Erosion of Cemented Carbide, *Proceedings of International Conference on Wear of Materials,* American Society of Mechanical Engineers, 1989, p 135-142

101. K. Anand, Effect of Microstructure on Local Impact Damage and Erosion of Cemented Carbides, Ph.D. dissertation, North Carolina State University, 1987

102. A.V. Levy, Ed., *Proceedings of Conference on Corrosion-Erosion-Wear of Materials at Elevated Temperatures,* NACE/EPRI/LBL/DOE-FE, Berkeley, CA, 31 Jan-2 Feb 1990

103. G. Sorrel, Elevated Temperature Erosion-Corrosion of Alloys in Sulfidizing Gas-Solids Streams: Parametric Studies, *Proceedings of Conference on Corrosion-Erosion-Wear of Materials at Elevated Temperatures,* A.V. Levy, Ed., NACE/EPRI/LBL/DOE-FE, Berkeley, CA, 27-29 Jan 1986, p 204-229

104. W.T. Bakker, Materials Performance in Coal Gasification Plants, *Proceedings of Conference on Corrosion-Erosion-Wear of Materials at Elevated Temperatures,* A.V. Levy, Ed., NACE/EPRI/LBL/DOE-FE, Berkeley, CA 27-29 Jan 1986, p 27-43

105. I.M. Hutchings, J.A. Little, and A.J. Ninham, Low Velocity Erosion-Corrosion of Steels in a Fluidized Bed, Paper 14, *Proceedings of Conference on Corrosion-Erosion-Wear of Materials at Elevated Temperatures,* A.V. Levy, Ed., NACE/EPRI/LBL/DOE-FE, Berkeley, CA, 31 Jan-2 Feb 1990

106. J. Zhou and S. Bahadur, Further Investigations on the Elevated Temperature Erosion-Corrosion of Stainless Steels, Paper 13, *Proceedings of Conference on Corrosion-Erosion-Wear of Materials at Elevated Temperatures,* A.V. Levy, Ed., NACE/EPRI/LBL/DOE-FE, Berkeley, CA, 31 Jan-2 Feb 1990

107. A.V. Levy, U.S. Dept. of Energy AR&TD Fossil Energy Materials Program, Quarterly Progress Report, Oak Ridge National Laboratory, 31 March 1986

108. V.K. Sethi and R.G. Corey, High Temperature Erosion of Alloys in Oxidizing Environments, Paper 73, *Proceedings of 7th International Conference on Erosion by Liquid and Solid Impact,* J.E. Field and J.P. Dear, Ed., Cambridge University Press, 1987

109. F.H. Stott, M.M. Stack, and G.C. Wood, The Role of Oxides in the Erosion-Corrosion of Alloys Under

Low Velocity Conditions, Paper 12, *Proceedings of Conference on Corrosion-Erosion-Wear of Materials at Elevated Temperatures,* A.V. Levy, Ed., NACE/EPRI/LBL/DOE-FE, Berkeley, CA, 31 Jan-2 Feb 1990

110. A.J. Ninham, I.M. Hutchings, and J.A. Little, in *Corrosion '89,* National Association of Corrosion Engineers, 1989
111. V.K. Sethi and I.G. Wright, Observations on the Erosion-Oxidation Behavior of Alloys, *Proceedings of Corrosion and Particle Erosion at High Temperature,* V. Srinivasan and K. Vedula, Ed., TMS-AIME, 1989, p 245-263
112. D.M. Rishel, F.S. Petit, and N. Birks, Some Principal Mechanisms in the Simultaneous Erosion and Corrosion Attack of Metals at High Temperature, Paper 16, *Proceedings of Conference on Corrosion-Erosion-Wear of Materials at Elevated Temperatures,* A.V. Levy, Ed., NACE/EPRI/LBL/DOE-FE, Berkeley, CA, 31 Jan-2 Feb 1990
113. C.T. Kang, F.S. Petit, and N. Birks, Mechanisms in the Simultaneous Erosion-Oxidation Attack of Nickel and Cobalt at High Temperature, *Metall. Trans.,* Vol 18A, 1987, p 1785-1803
114. C.T. Kang, S.L. Chang, F.S. Petit, and N. Birks, Synergism in the Degradation of Metals Exposed to Erosive High Temperature Oxidizing Temperatures, *Proceedings of Conference on Corrosion-Erosion-Wear of Materials at Elevated Temperatures,* A.V. Levy, Ed., NACE/EPRI/LBL/DOE-FE, Berkeley, CA, 27-29 Jan 1986, p 61-76
115. H.H. Uhlig, *Corrosion and Corrosion Control,* 2nd ed., John Wiley & Sons, 1971
116. S.L. Chang, F.S. Petit, and N. Birks, Effects of Angle of Incidence on the Combined Erosion-Oxidation Attack of Nickel and Cobalt, *Oxid. Met.,* 1989
117. D.M. Rishel, F.S. Petit, and N. Birks, The Erosion-Corrosion Behavior of Nickel in Mixed Oxidant Atmospheres, *Proceedings of Corrosion and Particle Erosion at High Temperature,* V. Srinivasan and K. Vedula, Ed., TMS-AIME, 1989, p 265-313
118. A.V. Levy, E. Slamovich, and N. Jee, Elevated Temperature Combined Erosion-Corrosion of Steels, *Wear,* Vol 110, 1986, p 117-149
119. S. van der Zwaag and J.E. Field, The Effect of Thin Hard Coatings on the Hertzian Field, *Philos, Mag.,* Vol 46, 1982, p 133-150
120. A.V. Levy, Mechanisms of Combined Erosion-Corrosion of Steels at Elevated Temperatures, *Proceedings of Corrosion and Particle Erosion at High Temperature,* V. Srinivasan and K. Vedula, Ed., TMS-AIME, 1989, p 207-230
121. V.K. Sethi and I.G. Wright, A Description of Erosion-Oxidation Based on Scale Removal and Scale Erosion, Paper 18, *Proceedings of Conference on Corrosion-Erosion-Wear of Materials at Elevated Temperatures,* A.V. Levy, Ed., NACE/EPRI/LBL/DOE-FE, Berkeley, CA, 31 Jan-2 Feb 1990
122. S. Hogmark, A. Hammarsten, and S. Soderberg, On the Combined Effects of Corrosion and Erosion, Paper 37, *Proceedings of 6th International Conference on Erosion by Liquid and Solid Impact,* J.E. Field and N.S. Corney, Ed., Cambridge University Press, 1983
123. G. Sundararajan, An Analysis of the Erosion-Oxidation Interaction, *Wear,* Vol 145, 1990, p 251-282
124. D.J. Stephenson, J.R. Nicholls, and P. Hancock, The Influence of Scale/Substrate Properties on High Temperature Erosion in Gas Turbines, Paper 48, *Proceedings of 6th International Conference on Erosion by Liquid and Solid Impact,* J.E. Field and N.S. Corney, Ed., Cambridge University Press, 1983
125. A.V. Levy and Y.F. Man, Elevated Temperature Erosion-Corrosion of 9Cr-1Mo Steel, *Wear,* Vol 111, 1986, p 135-159
126. A.V. Levy and Y.F. Man, The Effect of Temperature on the Erosion-Corrosion of 9Cr-1Mo Steel, *Wear,* Vol 111, 1986, p 161-172
127. A.V. Levy and Y.F. Man, Surface Degradation of Ductile Metals in Elevated Temperature Gas-Particle Streams, *Wear,* Vol 111, 1986, p 173-186
128. A.V. Levy and Y.F. Man, Erosion-Corrosion Mechanisms and Rates in Fe-Cr Steels, *Wear,* Vol 131, 1989, p 39-51
129. J. Zhou and S. Bahadur, High-Temperature Erosion-Corrosion Behavior of Stainless Steels, *Proceedings of Corrosion and Particle Erosion at High Temperature,* V. Srinivasan and K. Vedula, Ed., TMS-AIME,

1989, p 315-333

Cavitation Erosion

Carolyn M. Hansson, Queen's University, Canada; Inge L.H. Hansson, Alcan International Ltd., Canada

Introduction

CAVITATION is defined as the repeated nucleation, growth, and violent collapse of cavities, or bubbles, in a liquid. The term has been used in this context by the fluid mechanics and physics communities for centuries. It should not be confused with the metallurgical use of the term, which describes the growth of voids within a solid material, usually as a result of creep.

When a liquid is subjected to sufficiently high tensile stresses, vapor-filled voids, or cavities, are formed at weak regions within the liquid and usually grow under tensile conditions. In practice, all liquids contain gaseous, liquid, and solid impurities, which act as nucleation sites for the cavities. When the liquid that contains cavities is subsequently subjected to compressive stresses, that is, to higher hydrostatic pressures, these cavities will collapse. This collapse is directly responsible for the erosion process.

In practice, cavitation can occur in any liquid in which the pressure fluctuates either because of flow patterns or vibrations in the system. If, in a particular location in a liquid flow system, the local pressure falls below the vapor pressure of the liquid, then cavities may be nucleated, grow to a stable size, and be transported down-stream with the flow. When they reach a higher-pressure region, they become unstable and collapse, usually violently. This form of cavitation commonly occurs in hydrofoils, pipelines, hydraulic pumps, and valves. The pressures produced by the collapse may cause localized deformation and/or removal of material (erosion) from the surface of any solid in the vicinity of the cavities.

Similarly, when a stationary liquid is subjected to vibrational pressure fluctuations, the fluctuations may be sufficient to nucleate, grow, and collapse cavities, again resulting in erosion of any solid in the vicinity of the cavity cluster. Such cavities can produce the type of erosion that is typically observed on the coolant side of a diesel engine cylinder liner.

The collapse velocity, v, of a cavity is a function of the hydrostatic pressure, P, under which the cavity collapses, the volume, V, of the initial cavity, and the density, ρ, of the liquid (Ref 1):

$$v \propto \left(\frac{P\,V}{\rho}\right)^{1/2} \qquad \textbf{(Eq 1)}$$

Cavity radii for flow cavitation are typically from ~0.25 to 1.0 mm (~0.01 to 0.04 in.), and for vibratory cavitation, ~50 μm (~2 mils). P is of the order of a few atmospheres. For a cavity of 1 mm (0.04 in.) radius collapsing at 0.1 MPa (1 atm) overpressure in water, the cavity collapse velocity typically ranges from 100 to 150 m/s (330 to 490 ft/s).

Furthermore, the collapse time, t, of the cavity is related to the initial radius, R_0, of the cavity, the liquid density, ρ, and the hydrostatic pressure at collapse, P, as follows (Ref 1):

$$t \propto R_0\left(\frac{\rho}{P}\right)^{1/2} \qquad \textbf{(Eq 2)}$$

For the above case, the time is ~100 ns.

The effects of surface tension and viscosity on the collapse of a cavity are relatively insignificant. However, the compressibility of the liquid, vapor, and any trapped gases has a profound effect on the final stages of the collapse and will cushion the erosive effect of the single cavities (Ref 2). It is important to note that the driving force for the cavity collapse is the difference between the hydrostatic pressure and the vapor pressure of the liquid. A more detailed description of cavity dynamics and the parameters influencing the cavitation process is given by Mørch (Ref 3).

The mechanism by which cavitation causes erosion is briefly described below. When a cavity collapses within the body of the liquid, away from any solid boundary, it does so symmetrically and emits a shock wave into the surrounding liquid. On the other hand, those cavities that are either in contact with or very close to a solid surface will collapse asymmetrically, forming a microjet of liquid directed toward the solid, as shown in Fig. 1 (Ref 4).

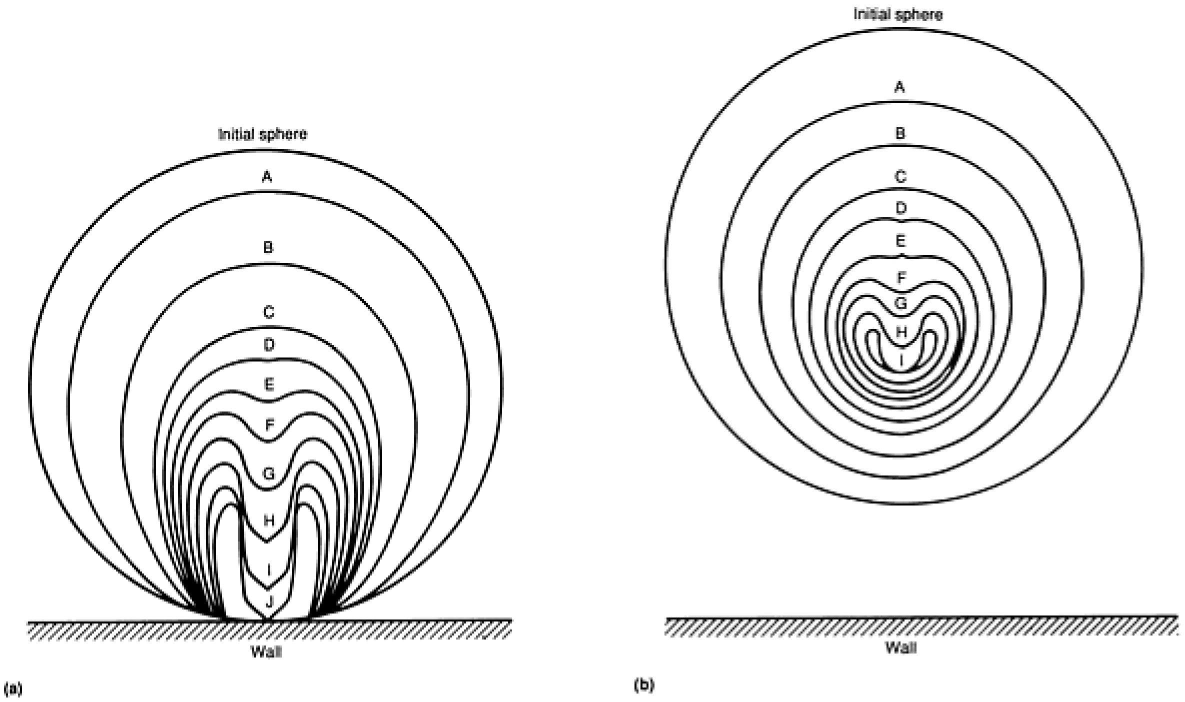


Fig. 1 Asymmetrical collapse of cavity. (a) In contact with solid surface. (b) Adjacent to solid surface. Source: Ref 4

The shock wave from spherical collapse and the jet impact from asymmetrical collapse have earlier been regarded as the most likely causes of erosion. However, each has features that do not permit a ready explanation of the observed erosion phenomena (Ref 5, 6). For example, the shock wave attenuates too rapidly, and the microjet diameters are typically too small to account for the degree and extent of the overall erosion damage. The discrepancies are now attributed (Ref 6, 7) to the fact that single cavities do not act independently but instead collapse in concert. The collapse of the cavity cluster enhances the effects of the cavities adjacent to, or in contact with, the solid.

In all practical situations involving cavitation, large numbers of cavities are generated at the same time and constitute what can be described as cavity clusters (Ref 2). When these clusters are subjected to an increased external hydrostatic pressure, they collapse in a concerted manner, stating with the cavities at the outer perimeter of the cluster and proceeding inward toward the central cavities (Ref 8). In this sequence, much of the energy generated by the collapse of the outer cavities is transferred to the cavities in the inner part of the cluster through an increased local hydrostatic pressure at the individual collapse. This results in a significant increase in the intensity of collapse of the central cavities (Ref 9). This concerted collapse mechanism has been demonstrated for flow cavitation as well as for vibratory cavitation (Ref 8, 10).

Because of the localized nature of the cavitation process, the energy dissipation has a significant temperature increase associated with the collapse. Local temperatures up to 5000 K have been reported (Ref 11).

Cavitation Erosion

Cavitation erosion is the mechanical degradation of materials caused by cavitation in liquids. The mechanical loading of a solid surface that is due to a cavitating liquid is caused by asymmetrical collapse of cavities either at or near the surface. These asymmetrical collapses result in liquid microjets that are directed toward the solid surface. The mechanical loads are very localized and, because of the concerted collapse of the cavity cluster, can be extremely severe, resulting in deformation of the surface. The repeated loading eventually leads to removal of material from the surface, that is, erosion.

The erosive effect of a cavity cluster is dependent on a number of factors, including hydrostatic pressure, cavity cluster size, distance of the individual cavities from the solid surface, cavity size distribution, and the temperature and density of the liquid. The total inherent energy of the cavity cluster is transferred to the solid material and must be either absorbed or dissipated by the solid or reflected as shock waves in the liquid. The solid material will absorb the impact energy as elastic deformation, plastic deformation, or fracture; the latter two processes lead to erosion of the material. The more elastic or plastic deformation energy that the material can absorb, the greater will be the cavitation erosion resistance of the material.

Erosion is generally regarded as mass loss from the surface and, for most materials under most forms of cavitation, is preceded by an incubation period, during which the material will deform either elastically or plastically (Ref 5). However, it should be noted that for some applications, a roughening of the surface by plastic deformation without actual loss of mass can render the part unusable for that application. Thus, cavitation damage without mass loss can be regarded as "erosion" for those applications. At the other extreme, material losses of >10 mm/year (0.4 in./year) from tough construction materials are observed in such applications as hydraulic turbines (Ref 12).

After the initiation of material loss from the surface, the rate of erosion as a function of continued exposure to cavitation is usually nonlinear (Ref 5). The observed time dependencies of the erosion rates are similar to those described in the article "Liquid Impingement Erosion" in this Volume.

Materials Factors

Localized Nature of Material Deformation and Removal. As described above, the loading that a material experiences during exposure to a cavitating liquid is localized, dynamic, and, at least initially, compressive in nature. However, the localized nature of the loading and the fact that it occurs at a free surface means that the deformation is not under the constraints normally imposed under bulk compressive or shock stressing. Therefore, the material is free to deform, both on a local and an extended level, in a manner that is unlike that of any other, more common, form of stressing. Because the material is not deformed as a whole, the deformation of one grain or one phase within a grain is not influenced by the behavior of the surrounding grains of phases, as it would be under bulk deformation conditions. Thus, the theoretical and empirical "rules" that have been developed to describe and explain the various strengthening mechanisms in different materials do not always apply to the resistance of the material to cavitation erosion.

Added to this localized loading factor is the dynamic, shocklike nature of the loading. It is not surprising, therefore, that no universal correlation with quasistatic mechanical properties has been observed. Consequently, the approach to materials selection and/or materials development for cavitation erosion resistance cannot be based on general experience in the selection and/or development of materials for resistance to bulk deformation and has, therefore, been almost exclusively empirical.

Until very recently, structural components have been predominantly fabricated of metallic alloys. A search of the literature has revealed very little information concerning cavitation erosion of either bulk ceramics of polymers. The only data concerning cavitation erosion of nonmetallic structures in practice has been of concrete (Ref 13), which, as a structural member of dams and sluices, is often exposed to cavitating liquids. Consequently, the discussion of the materials aspects of cavitation erosion will concentrate on metallic alloys and on the coatings and surface treatments that have been employed to minimize the erosion rates. There is little discussion of the very limited data on laboratory studies of nonmetallic bulk materials.

Erosion of Metals and Alloys. The deformation and failure mechanisms of both metals and alloys are markedly influenced by strain-rate sensitivity (and, therefore, the crystal structure) and the ability to absorb the energy of the shock loading without macroscopic deformation (which is related to the stacking fault energy). In multiphase alloys, the volume fraction, size, and dispersion of a second phase generally have a different and usually less significant influence on erosion rates than they do on the quasistatic mechanical properties.

Face-centered metals and alloys are isotropic and are the least sensitive to strain rate of the three common metallic structures. Consequently, their response to cavitation is similar to their quasistatic mechanical behavior in that they are highly ductile and fail by a void growth and coalescence mechanism (Ref 7) or by a ductile rupture (Ref 14) mechanism. Very early damage in the face-centered cubic (fcc) metals and single-phase fcc alloys consists of isolated depressions (Ref 15), Fig. 2, which can be attributed to the jet impact of individual cavities collapsing close to the surface. Also during this early stage, the grain boundaries become delineated, coarse slip bands develop across the width of the grains, and the grains become increasingly undulated. Eventually, the undulations develop into craters and material is lost by necking of the rims of the craters.

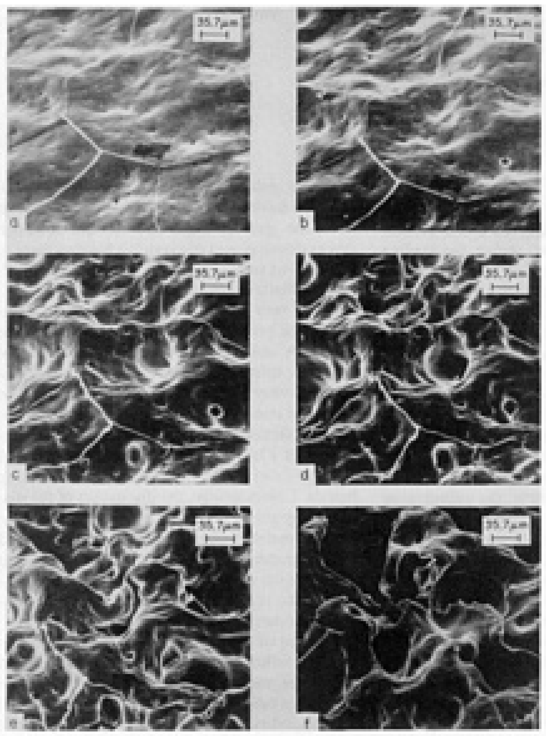


Fig. 2 Scanning electron micrographs of polycrystalline aluminum exposed to vibratory cavitation at varying lengths of time. (a) 12 s. (b) 24 s. (c) 40 s. (d) 60 s. (e) 75 s. (f) 90 s. Source: Ref 7

Body-centered cubic (bcc) metals and alloys are usually also isotropic, but their deformation is highly strain-rate sensitive. Their response to an applied stress is always a competition between flow and fracture. As the temperature decreases or the strain rate increases, flow becomes more difficult, and there is an increased tendency to brittle fracture. When pure iron is subjected to vibratory cavitation, it exhibits both brittle and ductile failure mechanisms. The brittle failure mode is illustrated in Fig. 3.

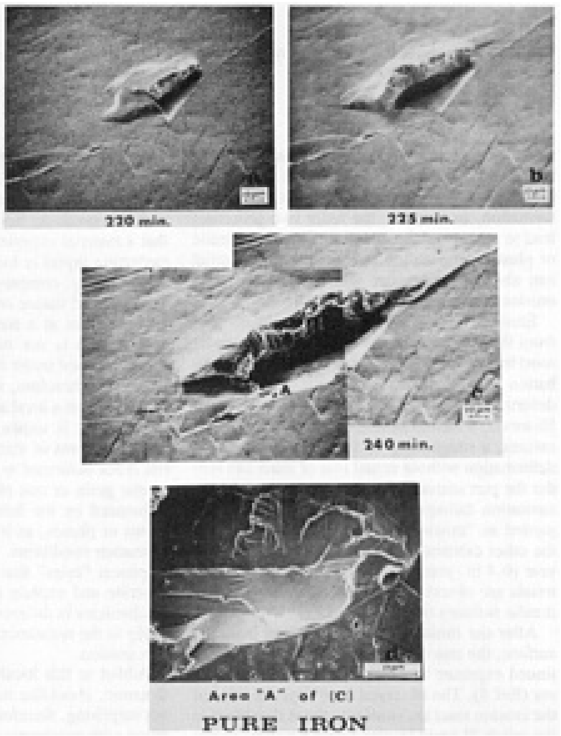


Fig. 3 Brittle fracture of pure iron exposed to vibratory cavitation. Source: Ref 16

Hexagonal close-packed (hcp) metals are anisotropic to varying degrees and can be either strain-rate sensitive or not, depending on the axial ratio of the unit cell. Two hcp metals that behave quite differently are zinc and cobalt. Zinc is highly anisotropic, highly strain-rate sensitive, and exhibits poor resistance to cavitation (Ref 17). Cobalt, on the other hand, is the most erosion-resistant pure metal of all those studied to date. Cobalt also has an almost ideal axial ratio and, unlike zinc, does not undergo a ductile-to-brittle transition with decreasing temperature or increasing strain rate. Its excellent erosion resistance is attributed to its extensive twinning mechanisms, which can effectively allow absorption of the cavitation energy without any major distortion of the metal (Ref 18).

Studies of multiphase alloys have shown that the size and dispersion of the second phase(s) determine whether or not these phases influence the cavitation erosion behavior. For example, the erosion rates and mechanism of material removal of precipitation-hardenable aluminum alloys exposed to cavitation are strongly dependent on the heat treatment, whereas the incubation period is little affected (Ref 19). Al-Mg alloys generally exhibit superior erosion resistance than do Al-Cu alloys because of the greater propensity for strain aging in the former. With both increasing solute content and degree of hardening, the mode of failure changes from ductile rupture, similar to that of the pure fcc metals, to the development of flat-bottomed pits that grow parallel to the surface and exhibit striated surfaces reminiscent of fatigue fracture surfaces. This effect appears to be related to the work hardenability of the surface layers and the depth of the work-hardened layers.

In steels, the ferrite phase controls the resistance to cavitation, because it erodes in a manner similar to pure iron. Thus, the microstructures with a continuous ferrite phase, for example, a spheroidized or normalized alloy, offer the least resistance because the ferrite can be eroded away from around the carbides, which then drop out (Ref 16, 20). It is necessary to strengthen the matrix phase, such as by heat treating to produce martensite or bainite, in order to impart greater erosion resistance.

In practical applications where intense cavitation is unavoidable, cobalt-base alloys (Ref 21) and, to a lesser extent, austenitic stainless steels (Ref 12) have been found to be the most erosion-resistant alloys available to date, despite their fairly low strength and hardness characteristics. Unlike most high-strength alloys, neither exhibits any significant strain-rate-sensitive behavior. More importantly, both have low stacking-fault energies and are readily able to develop stacking faults, twins, and/or martensitically transformed regions (Ref 18, 21, 22, 23). Thus, these alloys have the ability to absorb the impact energy with a lower distortion of the crystal lattice than do other alloys, which deform by normal multiple slip mechanisms.

The cavitation erosion resistance of cobalt alloys is superior to that of the stainless steels, but their cost is considerably higher. The economic factors, together with a better understanding of the factors responsible for the high erosion resistance of the cobalt alloys, have simulated the recent development (Ref 12) of new cavitation-resistant iron-base alloys. The success of this development is evidenced by the erosion data given in Fig. 4 for the new alloys designated IRECA and several iron- and cobalt-base commercial alloys. More comparative data on the erosion rates of various materials are given in the article "Liquid Impingement Erosion" in this Volume.

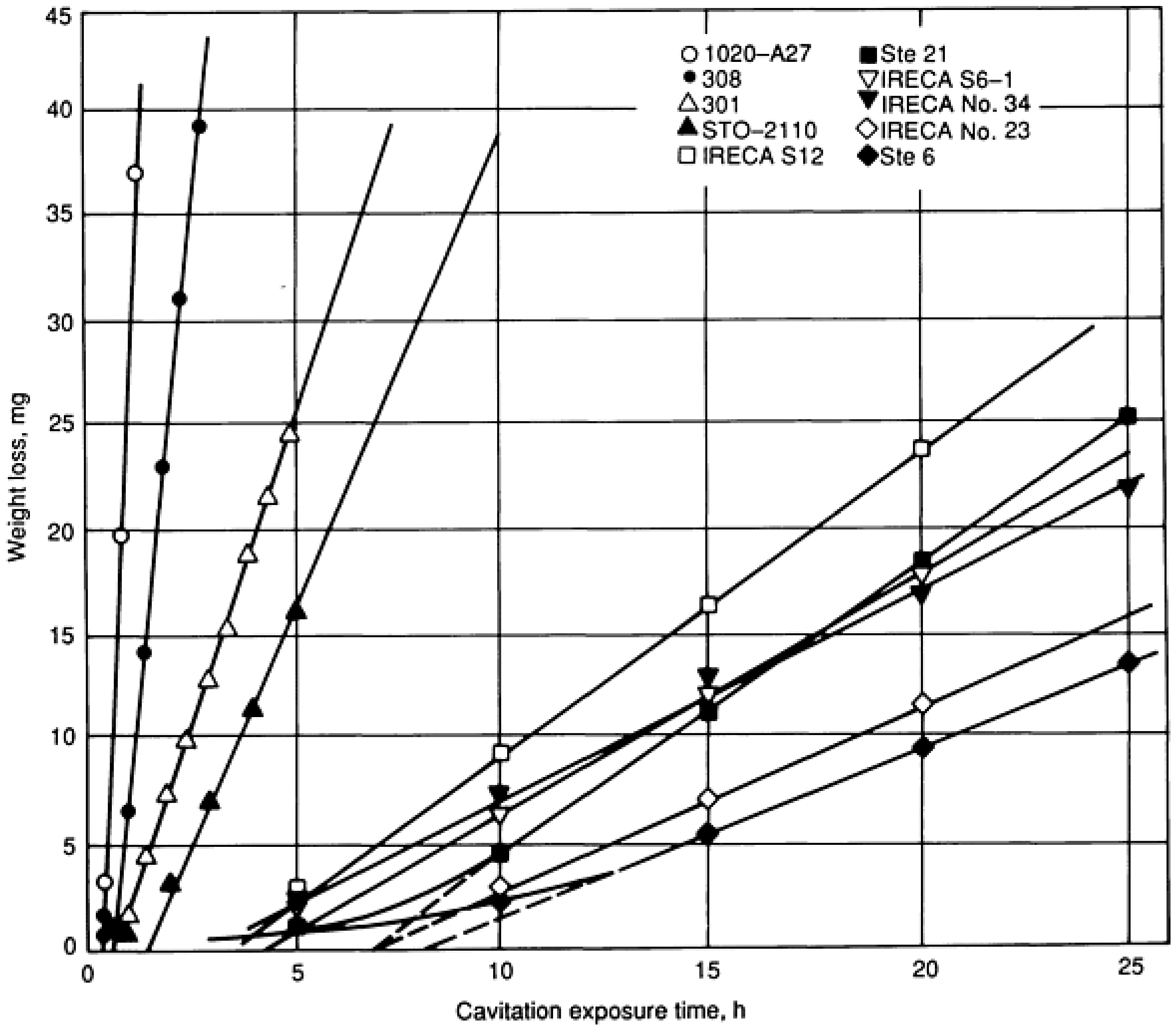


Fig. 4 Cavitation erosion weight loss as function of exposure time measured on various standard and experimental alloys in ASTM G 32 vibratory tests. Source: Ref 12

Surface Coatings and Treatments. The recent trend in materials development has been to specify a component for bulk property requirements and subsequently coat or treat the surface to provide the required environmental resistance. Welded overlays of stainless steels or cobalt alloys are commonly used to provide resistance to cavitation erosion. Several alternative coating techniques that have been investigated in the laboratory in recent years include: arc-sprayed coatings (Ref 24), plasma-sprayed coatings (Ref 25, 26), laser hardening, cladding and alloying (Ref 10, 27), ion implantation (Ref 28, 29), and electroless nickel coatings (Ref 30). Plasma-sprayed coatings and laser treatments are beginning to be applied in practice for this purpose.

Of these techniques, plasma spraying (Ref 31) is probably the most commercially well developed and offers flexibility in terms of the types of materials that can be sprayed. In addition, the technique can be carried out in air, in inert gas atmospheres, or in a reduced-pressure environment.

There are two major disadvantages of plasma-sprayed coatings. One is that the coating is generally only mechanically bonded to the substrate and therefore does not exhibit very good adhesion. The other disadvantage is the inherent porosity

of the coatings. One of the more erosion-resistant coatings produced by plasma spraying is of the "shape memory" alloy, NiTi (Ref 24). Like cobalt and the IRECA stainless steel, NiTi owes its superior properties to a stress-induced phase transformation.

Laser surface treatments appear to offer the most possibilities and advantages. By rapidly heating (but not melting) and quenching the surface of a finished component, an extremely hard and resistant surface layer can be induced without requiring any further surface finishing and with little effect on the bulk properties of the part. Alternatively, the composition and properties of the surface can be tailored specifically to the requirements either by melting the surface layers and adding additional alloy components to the base alloy or by depositing a cladding material onto the surface.

The advantages of this process are that the clad, alloyed, or heat-treated surface layer is an integral part of the component, which precludes any adhesion problems, and the process can be carried out in the atmosphere, rather than in vacuum. Its major disadvantages are that the processing equipment (laser and manipulation stations) is expensive and the technique cannot readily be executed on internal surfaces, because it is a line-of-sight process.

Laser cladding is now being used to provide erosion resistance to marine engine diesel cylinder liners (Ref 32). To the best of the authors' knowledge, the other laser techniques have not yet been applied for the purpose of cavitation erosion resistance.

Combined Effects of Cavitation Erosion and Corrosion

As described above, cavitation erosion leads to mechanical degradation of engineering materials, whereas corrosion is an electrochemical oxidation, or dissolution, of the material. Because cavitation always takes place in a liquid medium, there is always the possibility of an interaction between mechanical and electrochemical processes, which can produce diverse and complex effects on the materials. The interaction may be synergistic and can lead to increased damage. Alternatively, one mechanism may inhibit or reduce the harmful effects of the other, leading to a reduction in the overall damage.

Effect of Cavitation on the Corrosion Process. Cavitation can have a variety of effects on corrosion processes, including:

- Removing any protective passive film from the metal surface
- Increasing the diffusion rates of reactive dissolved gases to the metal surface
- Increasing the rate of removal of the corrosion reaction products from the vicinity of the surface

The net effect of cavitation is dependent on the type of corrosion. For example, it has been shown that cavitation can increase the ability of solution-treated stainless steel to become passive, whereas, for the same steel in the sensitized condition, the degree of intergranular corrosion is increased by cavitation (Ref 33).

Effect of Corrosion on the Cavitation Process. The corrosion process is electrochemical and can be described by two reactions: the anodic reaction, which involves the dissolution, or oxidation, of the metal, and the cathodic reaction, which usually involves the evolution of hydrogen. As mentioned above, dissolved gases can cushion the implosion of the cavities and reduce their damaging effects. In a situation that involves both corrosion and cavitation, the evolution of hydrogen can therefore have the effect of reducing the mechanical stressing of the metal. Similarly, it is possible (although no evidence has been reported) that solid particles produced by the corrosion process could act as nuclei for cavities, and thereby enhance the onset of cavitation.

Cavitation Erosion Testing

When testing materials for their cavitation erosion resistance, there is *no* laboratory experimental equipment that simulates the total situation for a real structural component exposed to cavitating liquids. However, there are a number of laboratory techniques and procedures that can be used to, at least reasonably, rank a series of selected materials on the basis of cavitation erosion resistance. The most commonly used techniques today are flow channels, vibratory (ultrasonic) systems, and cavitating jets, all of which can simulate accelerated cavitation erosion in most materials. However, it is important to note that most real situations involving cavitation also involve corrosion attack (for example, salt water on ship propellers) and other mechanical loading of the materials (for example, the structural load on valve seats or in concrete water channels).

Flow Channels. Typical flow channel equipment consists of a closed-loop circulating liquid flow channel with either a test section for scaled components, such as ship propellers, or a venturi restriction with a specimen holder designed to generate cavitation at specific locations near the specimen. This test simulates a flow cavitation situation very well. However, it is difficult to conduct accelerated cavitation erosion testing without changing the cavitation parameters relative to the service envelope of the simulated application. There are several different cavitation and specimen section designs (Ref 5) with the common feature that they are an integral part of the flow channel, which makes specimen changes difficult and more time consuming relative to the other techniques.

Vibratory (ultrasonic) equipment consists of an ultrasonic horn that is partly submerged in the liquid, which is contained in a beaker (Fig. 5). The vibration, typically at 20 kHz frequency, generates negative pressure for cavitation nucleation and growth, and positive pressure for cavity collapse in a small, stationary volume of the liquid. The specimen is either mounted on the horn tip (moving specimen) or at a fixed distance (a few millimeters) below the horn tip (stationary specimen).

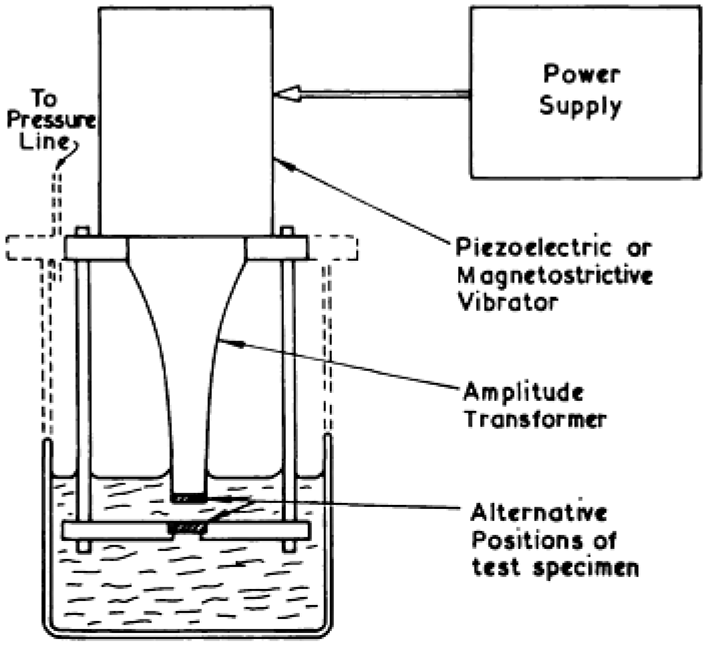


Fig. 5 Vibratory cavitation device in which specimen is either attached to or held below a horn oscillating in the lower kilohertz frequency range. Source: Ref 5

This test device is used for accelerated testing and lends itself to the study of interaction mechanisms with corrosion. Because the cavity size distribution is not the same as in the flow channel equipment, direct comparisons are not advisable. However, because the equipment is easy to use, it is widely applied to cavitation erosion resistance screening. Furthermore, ASTM G 32 describes the equipment and procedures for this test.

Cavitating Jet. Two variations of this technique have been described. In type I, a hydraulic pump with an accumulator delivers the test liquid through a sharp-entry parallel-bore nozzle, which discharges a jet of liquid into a chamber at a controlled pressure (Ref 34). In type II, a high-pressure nozzle with an internal center body is used to create the low pressure to initiate the cavitation (Ref 35, 36). Cavitation starts in the *vena contracta* region of the jet within the nozzle (type I) or at the end of the center body (type II) before ejecting as a cloud of cavities around the emerging jet (type I) or in the center of the jet (type II). The specimen is placed in the path of the jet at a specific stand-off distance from the nozzle tip. The cavities collapse on the specimen, thereby causing erosion of the test material.

The advantage of this technique is that it is an accelerated test method that offers the possibility of control and, thereby, allows the possibility of changing most of the cavitation parameters. Furthermore, the cavity size distribution resembles that of a real flow situation more than does that produced in the vibratory system. The technique is currently under consideration by ASTM as a standard test method.

Other Techniques. Rotating disk test equipment has been used in earlier cavitation studies (Ref 37, 38, 39, 40). Such equipment consists of a rotating disk with specimen holder and cavitation sites (that is, holes in the disk) submerged in the liquid. The liquid is kept relatively stable in a chamber where the disk is rotated. This technique is no longer in common use; however, it can be used, for example, to simulate cavitation occurring in pump impellers.

Means of Combating Erosion

Materials Selection and Development. It is clear from the current understanding of deformation mechanisms in metals and alloys, and from the dynamic and localized nature of cavitation loading, that materials selection for erosion resistance should be based on the ability of alloys to absorb the impact energy by a nondestructive strain mechanism, such as twinning, stacking-fault formation, or a stress-induced martensitic-type phase transformation. Unfortunately, most standard mechanical testing is quasistatic in nature, and most cavitation erosion testing comprises weight loss measurements without any determination of the mechanism of material loss. Consequently, mechanical property databases do not usually contain the type of information necessary for appropriate materials selection.

Similarly, there has been little effort to develop materials specifically for their erosion-resistant properties. However, the success of the initial research into the development of the iron-base IRECA alloys, which was based on a knowledge of the required deformation mechanisms and understanding of the compositional factors necessary to ensure that the alloy could deform in the appropriate manner, suggests that this is indeed a feasible approach and should be pursued.

Coatings and Surface Treatments. Coating technology is one of the more rapidly growing technologies in the field of materials. It is clear that the selection of the base material for its bulk properties and a coating/surface treatment for its resistance to environmental factors is the wave of the future. A combination of the development of materials specifically designed for erosion resistance and the appropriate technique for the application of these materials as a coating would be the optimum solution. Suitable coating techniques also allow for regeneration of parts that have been rendered unusable by erosion.

Other measures described below include design, air injection, and control of the operating temperature or pressure.

System design represents the best way to either reduce or eliminate cavitation erosion. Therefore, fluid flow systems should be designed to minimize the changes in flow pressure that occur when the velocity is either increased or decreased, usually as a result of constrictions or changes in the direction of the flow. Similarly, the elimination of vibrations or the reduction in their amplitude would reduce the problems of cavitation erosion in many types of machinery. If cavitation cannot be eliminated, the cavitating regions should be designed to allow the cavities to collapse as far away from a solid surface as possible or to decrease the concerted collapse mode of the cavity cluster (Ref 9).

Air injection into a cavitating fluid has been shown to be an effective method of reducing the intensity of erosion. The air creates bubbles and partly fills the cavities as they are formed, which prevents their complete collapse, thereby significantly reducing the magnitude of the shock wave emitted or the impact pressure of the microjets. The air also significantly changes the dynamic properties of the liquid, that is, the shock-wave velocity and its attenuation.

Control of Operating Temperature or Pressure. It is more difficult to nucleate cavities at temperatures that approach the freezing temperature, T_F, of the liquid and more difficult to collapse them at temperatures that approach the boiling temperature, T_B. Therefore, cavitation erosion intensity is at a maximum at temperatures in the middle range between freezing and boiling. Consequently, a change in temperature that is close to either T_F or T_B will reduce the cavitation intensity and, thus, the degree of erosion. Similarly, increasing the hydrostatic pressure makes nucleation of cavities more difficult, but increases the erosive power of the cavities, whereas decreasing the collapse pressure makes collapse less intense.

More comprehensive discussions of the influence of various cavitation parameters on the resulting erosion are given in a number of reviews (Ref 5, 41, 42).

The salient conclusions that can be made about cavitation erosion are:

- No materials are immune to cavitation erosion, as some are to corrosion; all will eventually erode
- Metallic materials that exhibit stress-induced phase transformations have the highest erosion resistance. Further development of alloys specifically for their erosion resistance is to be encouraged

- The combination of erosion and corrosion can be either synergistic or less harmful than either process alone. Unfortunately, there do not appear to be any general rules; therefore, each combination of material, environment, and erosion conditions must be evaluated
- Coating technologies, particularly laser processing, offer great potential both in providing tailor-made erosion resistance to structures that are selected for their bulk properties and in repairing and regenerating eroded surfaces

References

1. Lord Rayleigh, *Philos. Mag.,* Vol 34, 1917, p 94-98
2. L. van Wijngaarden, *11th International Congress of Applied Mechanics* (Munich), 1964
3. K.A. Mørch, Dynamics of Cavitation Bubbles and Cavitating Liquids, *Erosion,* C.M. Preece, Ed., Academic Press, 1979, p 309-355
4. M.S. Plesset and R.B. Chapman, *J. Fluid Mech.,* Vol 47, 1971, p 283-290
5. C.M. Preece, Cavitation Erosion, *Erosion,* C.M. Preece, Ed., Academic Press, 1979, p 249
6. C.M. Preece and I.L.H. Hansson, A Metallurgical Approach to Cavitation Erosion, *Advances in the Mechanics and Physics of Surfaces,* R.M. Latanision and R.J. Courtel, Ed., Harwood Academic Publishers, 1981, p 191-253
7. B. Vyas and C.M. Preece, Cavitation-Induced Deformation of Aluminum, *Erosion, Wear and Interfaces with Corrosion,* ASTM, 1973
8. I.L.H. Hansson and K.A. Mørch, Comparison of the Initial Stage of Vibratory and Flow Cavitation Erosion, *5th International Conference on Erosion by Liquid and Solid Impact,* Cavendish Laboratory, Cambridge University, 1979
9. I.L.H. Hansson and K.A. Mørch, The Influence of Cavitation Guide Vanes on the Collapse of Cavity Clusters and on the Resulting Erosion, *11th Symposium of the IAHR Symposium on Operating Problems of Pump Stations and Power Plants* (Amsterdam), International Association for Hydraulic Research, 1982
10. C.M. Preece and C.W. Draper, The Effect of Laser Quenching the Surfaces of Steels on Their Cavitation Erosion Resistance, *Wear,* Vol 67, 1981, p 321
11. E.B. Flint and K.S. Suslick, The Temperature of Cavitation, *Science,* Vol 253, 1991, p 1397-1399
12. R. Simoneau *et al.,* Cavitation Erosion and Deformation Mechanisms of Ni and Co Austenitic Stainless Steels, *7th International Conference on Erosion by Liquid and Solid Impact,* Cavendish Laboratory, Cambridge University, 1987
13. M.J. Kenn and A.D. Garrod, Cavitation Damage and the Tarbela Tunnel Collapse of 1974, *Proc. Inst. Civil Eng.,* Vol 70, 1981, p 65
14. C.M. Preece, S. Vaidya, and S. Dakshinamoorthy, The Influence of Crystal Structure on the Response of Metals to Cavitation, *Erosion: Prevention and Useful Applications,* ASTM, 1979
15. I.L.H. Hansson and K.A. Mørch, The Initial Stage of Cavitation Erosion of Aluminum in Water Flow, *J. Phys.,* Vol D11, 1978, p 147-154
16. E.H.R. Wade and C.M. Preece, Cavitation Erosion of Iron and Steel, *Metall. Trans.,* Vol 9A, 1978, p 1299-1310
17. S. Vaidya and C.M. Preece, Cavitation-Induced Multiple Slip, Twinning, and Fracture Modes in Zinc, *Scr. Metall.,* Vol 11, 1977, p 1143-1146
18. S. Vaidya, S. Mahajon, and C.M. Preece, The Role of Twinning in the Cavitation Erosion of Cobalt Single Crystals, *Metall. Trans.,* Vol 11A, 1980, p 1139-1150
19. S. Vaidya and C.M. Preece, Cavitation Erosion of Age-Hardenable Aluminum Alloys, *Metall. Trans.,* Vol 9A, 1978, p 299-307
20. R. Schulmeister, *Proceedings of the 1st International Conference on Rain Erosion,* Royal Aircraft Establishment, United Kingdom, 1965
21. D.A. Woodford, *Metall. Trans.,* Vol 3, 1978, p 1137

22. J.W. Tichler and A.W.J.D. Gee, *3rd International Conference on Rain Erosion,* Royal Aircraft Establishment, United Kingdom, 1974

23. T.F. Pedersen, S. Pedersen, and I.L.H. Hansson, Subsurface Deformation Studies of Cavitation Eroded FCC Materials, *6th International Conference on Erosion by Liquid and Solid Impact,* Cavendish Laboratory, Cambridge University, 1983

24. A.P. Jardine, Y. Horan, and H. Herman, Cavitation Erosion Resistance of Thick Film Thermally Sprayed NiTi, *Proceedings of Symposia on High Temperature Intermetallics,* Vol 213, Materials Research Society, 1991, p 815-820

25. X.-X. Guo, H. Herman, and S. Rangaswamy, Cavitation Erosion of Plasma Sprayed WC/Co, *Advances in Thermal Spraying,* Pergamon Press, 1986, p 37-41

26. S. Sampath, G.A. Bancke, and H.R. Herman, Plasma Sprayed Ni-Al Coatings, *Surf. Eng.,* Vol 5 (No. 4), 1989, p 293-298

27. R.J. Crisci, C.W. Draper, and C.M. Preece, Cavitation Erosion Resistance of Laser Surface Melted Self-Quenched Fe-Al Bronze, *Appl. Opt.,* Vol 21, 1982, p 1730

28. W.W. Hu, *et al.,* Cavitation Erosion of Ion-Implanted 1018 Steel, *Mater. Sci. Eng.,* Vol 45, 1980, p 263-268

29. C.M. Preece and E.N. Kaufmann, The Effect of Boron Implantation on the Cavitation Erosion Resistance of Copper and Nickel, *Corros. Sci.,* Vol 22, 1982, p 267-281

30. S. Pedersen and I.L.H. Hansson, Nickel Coatings for Cavitation Erosion Resistance of Brass Components, *6th International Conference on Erosion by Liquid and Solid Impact,* Cavendish Laboratory, Cambridge University, 1983

31. H. Herman, Plasma-Sprayed Coatings, *Sci. Am.,* Vol 256 (No. 9), 1988, p 112-117

32. W. Amende, private communication, 1989

33. B. Vyas and I.L.H. Hansson, The Cavitation Erosion-Corrosion of Stainless Steel, *Corros. Sci.,* Vol 30 (No. 8/9), 1990, p 761-770

34. A. Lichtarowicz and P.J. Scott, Erosion Testing with Cavitating Jet, *5th International Conference on Erosion by Liquid and Solid Impact,* Cavendish Laboratory, Cambridge University, 1979

35. P.A. March, Evaluating the Relative Resistance of Materials to Cavitation Erosion: A Comparison of Cavitation Jet Results and Vibratory Results, *Cavitation and Multiphase Flow Forum,* FED, 1987

36. P.A. March, Cavitating Jet Facility for Cavitation Erosion Research, *Symposium on Cavitation Research Facilities and Techniques,* American Society of Mechanical Engineers, 1987

37. J.Z. Lichtman, D.H. Kallas, C.K. Chatten, and E.P. Cochran, Cavitation Erosion Resistance of Structural Materials and Coatings, *Corrosion,* Vol 17, 1961, p 497-505

38. J.Z. Lichtman and E.R. Weingram, The Use of a Rotating Disc Apparatus in Determining Cavitation Erosion Resistance of Materials, *Symposium on Cavitation Research Facilities and Techniques,* American Society of Mechanical Engineers, 1964

39. A. Thiruvengadam, A Comparative Evaluation of Cavitation Damage Test Devices, *Cavitation Research Facilities and Techniques,* American Society of Mechanical Engineers, 1964

40. P. Veerabhadra Rao, Correlating Models and Prediction of Erosion Resistance to Cavitation and Drop Impact, *J. Test. Eval.,* 1976, p 3-14

41. H. Wiegand and R. Shulmeister, Investigations with a Vibratory Apparatus on the Influence of Frequency, Amplitude, Pressure, and Temperature on Material Destruction by Cavitation, *Motortechnische Zeitschrift,* Vol 29 (No. 2), 1968, p 41-50

42. S. Pedersen, "Cavitation Erosion: Vibratory Cavitation and Cavitation Erosion of Metals," Ph.D. thesis, Laboratory of Applied Physics, Technical University of Denmark, 1986

Liquid Impingement Erosion

Frank J. Heymann (retired), Westinghouse Electric Corporation

Introduction

LIQUID IMPINGEMENT EROSION has been defined as "progressive loss of original material from a solid surface due to continued exposure to impacts by liquid drops or jets" (Ref 1). The operative words in this definition are "*impacts* by liquid drops or jets": liquid impingement erosion connotes repeated impacts or collisions between the surface being eroded and small discrete liquid bodies.

Excluded from this definition are erosion mechanisms due to the impingement of a continuous jet, due to the flow of a single-phase liquid over or against a surface, due to a cavitating flow, or due to a jet or flow containing solid particles--although all these can produce erosion (progressive loss of solid material) at least under some conditions. Some of these mechanisms will, however, be discussed briefly in order to distinguish them clearly from the primary subject.

The significance of the discrete impacts is that they generate impulsive contact pressures on the solid target, far higher than those produced by steady flows (see the discussion "Liquid/Solid Interaction--Impact Pressures" later in this article). Thus, the endurance limit and even the yield strength of the target material can easily be exceeded, thereby causing damage by purely mechanical interactions. In some circumstances the damage can also be accelerated by conjoint chemical action.

At sufficiently high impact velocities, solid material can be removed even by a single droplet (or other small liquid body). Much of what is currently known about the liquid/solid interactions in liquid impingement has been determined through laboratory experiments and analytical modeling involving single impacts.

Liquid impingement erosion in its advanced stages is characterized by a surface that appears jagged, composed of sharp peaks and pits (Fig. 1). A possible reason for this will be given later.

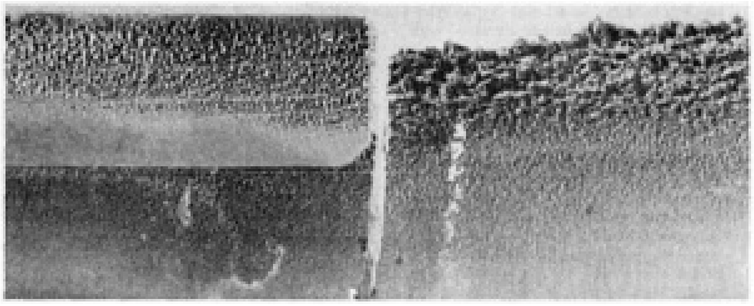


Fig. 1 Two portions of a steam turbine blade that has experienced liquid impingement erosion. The portion on the left was protected by a shield of rolled Stellite 6B brazed onto the leading edge of the blade; the portion on the right is unprotected type 403 stainless steel. Note the difference in degree of erosion. Normally such erosion does not impair the blade's function. Both at 2.5×

A very comprehensive treatment and review of liquid impact erosion can be found in Ref 2; in particular the chapters therein by Adler (Ref 3) and by Brunton and Rochester (Ref 4). Reference 5 contains some now classic studies that provided the foundation for subsequent work. Many other contributions to this field are found in several ASTM symposium volumes (Ref 6, 7, 8, 9, 10) and in the proceedings of the international "Rain Erosion" and "Erosion by Liquid and Solid Impact" (or "ELSI") conferences (Ref 11, 12, 13, 14, 15, 16, 17). Individual papers from some will be cited in context.

Acknowledgements

The author would like to thank John E. Field of Cambridge University, George F. Schmitt, Jr. of the Air Force Materials Laboratory and Westinghouse Electric Corporation for supplying photographs. Additional thanks are due to George Schmitt for also supplying information on the current state of rain erosion protection and providing valuable suggestions for improving this article.

Occurrences in Practice

It is quite difficult to propel liquid droplets to high velocities without breaking them up, and liquid impingement erosion haas become a practical problem primarily where the target body moves at high speeds and collides with liquid drops that are moving much more slowly. Almost all the work done in this subject has been in connection with just two major problems: "moisture erosion" of low-pressure steam turbine blades operating with wet steam, and "rain erosion" of aircraft or missile surfaces and helicopter rotors.

Whenever vapor or gas flows carrying liquid droplets impinge upon solid surfaces--as in nuclear power plant pipes and heat exchangers, for example--erosion can also occur. However, the probable impact velocities and impact angles are such as to make "pure" liquid impingement erosion an unlikely mechanism. It is much more likely that an "erosion-corrosion" mechanism is then involved (see the discussion of "Impingement Attack and Erosion-Corrosion" later in this article).

Steam Turbine Blade Erosion. Moisture erosion of low-pressure blades has been a problem throughout steam turbine history, and remains a concern today. In the last stages of the low-pressure turbine, the steam expands to well below saturation conditions, and a portion of the vapor condenses into liquid. Although the condensation droplets are very small, some of them are deposited onto surfaces of the stationary blades (guide vanes), where they coalesce into films or rivulets and migrate to the trailing edge. Here they are torn off by the steam flow, in the form of much larger droplets.

These large droplets slowly accelerate under the forces of the steam acting on them, and when they are carried into the plane of rotation of the rotating blades, they have reached only a fraction of the steam velocity. As a result, the blades hit them with a velocity that is almost equal to the circumferential velocity (wheel speed) of the blades, which can be as high as 650 m/s (2100 ft/s) in a modern 3600 rpm turbine. References 18 and 19 describe these processes in detail. The same basic phenomenon can, of course, occur in wet vapor turbines operating with other working fluids, such as sodium or mercury.

The principal remedies in modern turbines include extracting moisture between blade rows, increasing axial spacing between stator and rotor to permit droplets to be accelerated and broken up, and making the leading edge of the blade more resistant to erosion. This last remedy has been accomplished by local flame hardening of the blade material, by brazed-on "shields" of Stellite (Fig. 1), or in some cases by shields of tool steel or weld-deposited hardfacing. Tests on many blade and shield materials are reported in Ref 20 and 21. The base material for present-day low-pressure blades is usually a 12% Cr martensitic stainless steel, a 17Cr-4Ni precipitation-hardening stainless steel, or, more rarely, a titanium alloy.

Recently, success has been claimed for new "self-shielding" blade alloys that harden under the action of the impacts. One such alloy is Jethete M152, a martensitic steel containing about 11% Cr, 2.9% Ni, 1.6% Mo, and 0.3% V. Other new approaches that have been investigated include plasma-deposited Stellite and an ion-plated chromium-tin multilayer coating; however, it is doubtful that relatively thin coatings can provide long-term protection.

The evaluation and prediction of steam turbine blade erosion is very complex; recent contributions include Ref 22 and 23.

Aircraft Rain Erosion. Rain erosion became a major problem in the 1950s, when military aircraft reached transonic and supersonic speeds. The impact of rain drops, 2 mm (0.08 in.) or more in size, on unprotected aluminum alloy surfaces, optical and infrared windows, and radomes caused severe erosion which seriously limited operational time in rain storms. This resulted in many government-funded research projects into erosion mechanisms as well as development and evaluation of protective coatings. Reference 24 gives an overview of the rain erosion problem, with special reference to radomes. The current status concerning remedies has been summarized as follows by Schmitt (Ref 25):

Protection of aircraft radomes and composite surfaces is accomplished with two classes of elastomeric coatings--polyurethanes for widespread lower temperature applications, and fluorocarbons where elevated temperatures (above 177 °C, 350 °F) or special requirements (camouflage colors, thermal flash protection) are involved.

For applications where supersonic rain erosion is a concern, the inherent erosion resistance of the base materials combined with streamline geometry to reduce impact angles is the most often used approach. Another aerodynamic technique is to utilize the shock waves to shatter and fragment the raindrops into very small pieces that produce less damage. For hemispherical domes where impact angles must be large (near normal), protective coatings of boron phosphide, germanium carbon and diamond are being pursued, but they are restricted to velocities less than Mach 2. At extremely high velocities, protection at high impact angles may require metal tips or sacrificial layers even though a performance penalty must be paid. In many cases, lack of adequate materials and potential catastrophic failure simply precludes operation at high supersonic speed in rainy environments.

References 8, 9, 10, 11, 12, 13, 14, 15, 16, and 17 contain numerous papers on evaluation of materials and coatings for rain erosion applications. Among the most recent are investigations of infrared window materials, slip-cast fused silica, and hard carbon-coated germanium in Ref 17; polyurethane and fluoroelastomer coated composite constructions, composite materials for radomes, new materials for radomes, and infrared windows (including polyethersulfone, polyetherimide, polyetherketone, and germanium) in Ref 16; and slip-cast fused silica, boron-aluminum composites, composite and honeycomb structures, polytetrafluoroethylene (PTFE) and polymethylmethacrylate (PMM) in Ref 15.

As mentioned earlier, rain erosion also poses a threat to missile surfaces and helicopter rotors. Figure 2 shows the catastrophic failure of a missile dome due to rain erosion effects.

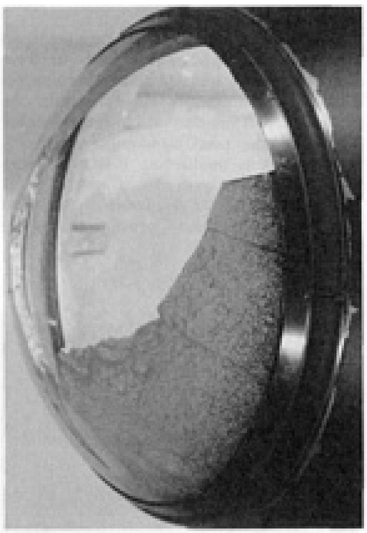


Fig. 2 Rain erosion effects on a Maverick missile dome made of coated zinc sulfide that was exposed for 10 s at a speed of about 210 m/s (690 ft/s). The dome itself suffered catastrophic damage, and erosion is also seen on the filled elastomeric mounting ring. Courtesy of G.F. Schmitt, Jr., Materials Laboratory, Wright Research and Development Center, Department of the Air Force

Relationship to Other Erosion Processes

Continuous Jet Impingement. Impingement of a high-velocity continuous jet can cause material removal, and that fact has led to the development of jet-cutting technology used in quarrying, mining, and material cutting. While there is some overlapping with erosion research, much of the literature is found in the "American Water Jet Conferences" and the "International Conferences on Jet Cutting Technology."

Steady continuous jet impingement produces only stagnation pressure on the target, whereas discrete impacts produce much higher shock-wave or "water-hammer" pressures. Thus, high supply pressures are needed to achieve material removal with truly continuous jets. This has led jet-cutting research to develop techniques for producing moving and oscillating jets, pulsating jets, and cavitating jets, all of which serve to introduce something akin to discrete impact conditions (also, jets laden with abrasive particles are used). References 26 and 27 describe these techniques.

Rough comparisons of test data suggest that the erosion rate due to a continuous jet can be from one to five orders of magnitude lower than that due to the same quantity of liquid impinging at the same velocity but in the form of droplets.

Cavitation Erosion. Whereas liquid impingement connotes a continuous vaporous or gaseous phase containing discrete liquid droplets, cavitation connotes a continuous liquid phase containing discrete vaporous or gaseous bubbles or cavities. Despite this seeming antithesis, the nature of cavitation damage and liquid impingement damage has many similarities. Both, in fact, are due to small-scale liquid/solid impacts. In cavitation, micro-jet impacts have been shown to occur in the asymmetrical collapse of cavitation bubbles adjacent to a solid surface, although shock waves generated by collapsing cavity clusters may also contribute to damage.

The relative resistance of materials to the two types of erosion is much the same, the damage appearance is similar, the complicated time dependence of the erosion rate is similar (see the section "Time Dependence of Erosion Rate" in this article), and historically cavitation tests have been used to screen materials for service in liquid impingement environments, and vice versa. In some practical cases, it is not clear whether the mechanism causing erosion was impingement or cavitation erosion. One such example is the heavily eroded dynamometer stator shown in Fig. 3. Erosion in such machines has often been characterized as cavitation erosion; however, the author believes it to be liquid impact erosion caused by the discrete streams of water issuing from the rotor pockets and sweeping across the stator vanes. This is supported by dynamic pressure transducer spectra which were dominated by discrete spikes at the rotor vane passing frequency, but did not show the characteristic signature of cavitation. In addition, injection of air bubbles, which can inhibit cavitation, had no influence on the signals or on the erosion.

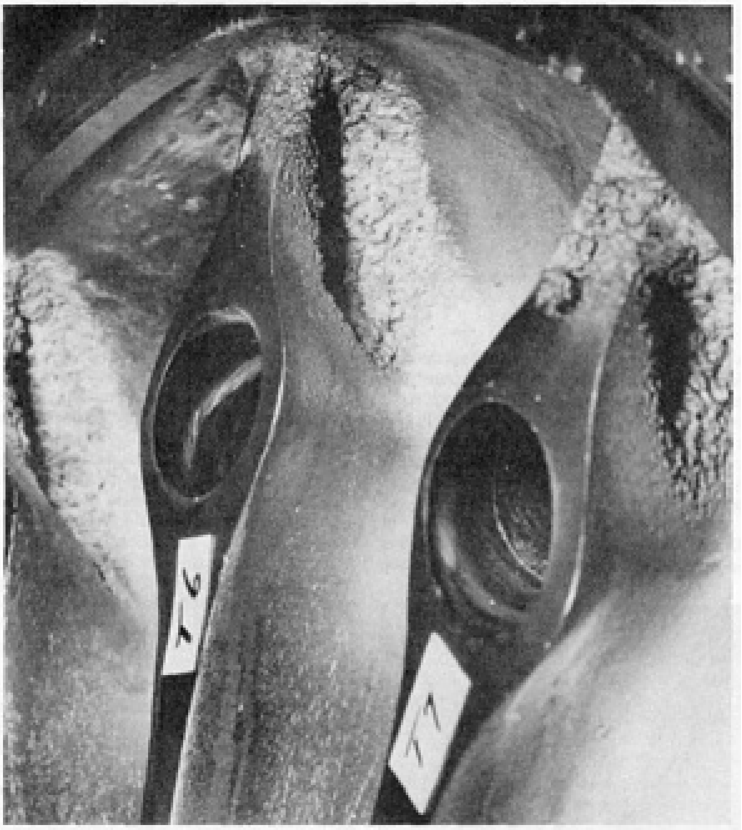


Fig. 3 Severe erosion of copper-manganese-aluminum stator vanes in a hydraulic dynamometer. Each vane has lost about 10 cm^3 (0.6 $in.^3$) of material. See text for a discussion of the erosion mechanism.

For additional information on cavitation damage, see the article "Cavitation Erosion" in this Volume.

Solid Particle Erosion. It might easily be assumed that solid particle erosion would have many similarities to liquid impingement erosion, since both involve the impact of small discrete bodies. This is not the case, however, because the damage mechanisms, the effects of impact variables, and the response of materials are all quite different.

For example, the resistance of common engineering alloys to liquid impingement ranges over several orders of magnitude, whereas most common ductile alloys have about the same resistance to solid impingement; liquid impingement erosion rates vary with about the 5th power of impact velocity, whereas solid particle erosion varies with about the 2.5th power; liquid impingement erosion is greatest with normal (perpendicular) impacts, whereas solid particle erosion in ductile alloys peaks with impacts some 60 to 70° away from perpendicular; and finally, liquid impingement erosion exhibits a complicated time dependence, whereas the erosion rate due to solid particle impingement is essentially linear. See the article "Solid Particle Erosion" in this Volume for further details.

Impingement Attack and Erosion-Corrosion. There are many practical situations where material loss occurs in the presence of flowing fluids but cannot be fully explained by purely mechanical action such as liquid impingement, solid particle impingement, or cavitation. Only recently has a substantial amount of scientifically oriented work on these phenomena been reported in the literature.

"Impingement attack" is a term sometimes used for material loss in tube bends and heat exchanger tube entrances, where the forces of unsteady, turbulent, or bubbly flows are believed to remove protective oxide layers and thus permit continuing and accelerated corrosion. "Wire drawing" is a term used for a grooving type of erosion produced in small gaps such as valve seats and component joints with a high pressure drop across them. This type of erosion could be due to localized cavitation or to erosion-corrosion.

"Erosion-corrosion" can refer to any conjoint (synergistic) action between corrosive and erosive processes, such that the resulting material loss rates are greater than the sum of the individual processes taken by themselves. Corrosive actions in the presence of solid particles, slurries, and sliding wear are covered in detail in the articles "Solid Particle Erosion," "Slurry Erosion," and "Corrosive Wear" in this Volume.

For purely fluid systems, some recent work on "corrosive-erosion" has emphasized a mechanism in which impulsive mechanical interactions play little or no role. For example, the flow of liquid or droplets along a carbon or low-alloy steel surface prevents equilibrium in the corrosive process and results in continuing chemical dissolution of the protective oxide (magnetite) layer (Ref 28). This is more accurately called "flow-assisted corrosion."

Only a few references can be cited here. Keck and Griffith (Ref 29) propose models for both convection-assisted oxide dissolution and for oxide fatigue by liquid droplet impacts. Coulon (Ref 30) distinguishes mechanisms by the flow velocity causing them: corrosion, 0 to 10 m/s (0 to 35 ft/s); corrosion-erosion, 10 to 50 m/s (35 to 165 ft/s); erosion-corrosion, 50 to 200 m/s (165 to 655 ft/s); and erosion, >200 m/s (>655 ft/s). Henzel *et al.* (Ref 31) give an update on experience and remedies. Both Ref 30 and 31 provide empirical schemes for estimating material loss due to erosion-corrosion, including factors relating to material, flow velocity, and temperature, as well as geometry factors borrowed from another empirical model by Keller (Ref 32).

Mechanisms of Liquid Impact Erosion

Liquid/Solid Interactions--Impact Pressures. The high-velocity impact of a liquid drop against a plane solid surface produces two effects that result in damage to that surface: high contact pressure, which is generated in the area of the impact, and subsequent liquid "jetting" flow along the surface, radiating out from the impact area (Ref 33). A first approximation of the average impact pressure, before radial outflow initiates, is the one-dimensional water-hammer pressure; that is, pressure generated in the impact of an infinite flat liquid surface against an infinite flat rigid surface. In this case a plane shock wave is formed at the instant of impact and travels into the liquid, bringing to rest one "layer" after another.

This impact or shock pressure, P, can be defined as:

$$P = \rho CV$$

where ρ is the liquid density, C is the shock wave velocity in the liquid, and V is the impact velocity. For practical impact velocities, this can be approximated by:

$$P = \rho C_0 V (1 + kV/C_0)$$

where C_0 is the acoustic velocity of the liquid, and $k = 2$ for water. For example, for water impacting at 500 m/s (1640 ft/s), this pressure is about 1250 MPa (180 ksi)--considerably above the yield strength of many alloys. The stagnation pressure of a continuous jet ($\rho V^2/2$) at that speed is about one-tenth of the former.

The real situation is much more complicated, because of the roundness of the impacting droplet and the elastic and plastic deformations of the solid surface. Although much work has been done on this topic in recent years (for example, Ref 34, 35, 36, 37), a complete understanding has not yet been achieved. However, the following salient features (see Fig. 4) are now widely accepted for the impact of a round drop on a rigid surface:

- At the initial instant of impact, contact is made at a point. Because the droplet radius of curvature is initially infinite compared to the contact radius, quasi-one-dimensional conditions prevail and a shock wave of the one-dimensional impact pressure is formed at that instant and begins to travel up into the drop (Fig. 4a)
- As the contact area spreads out, its perimeter at first is moving radially outward at a speed greater than the shock velocity; consequently an obliquely attached bulbous shock front is formed enclosing the compressed liquid and no free outflow can take place (Fig. 4b). During this stage, the highest impact pressure is found at the growing contact perimeter, where its value gradually increases up to about $3\rho C_0 V$, while that at the impact center decreases
- At some point, defined by a critical contact angle, ϕ_c, the conditions for an attached shock are no longer met; the shock front then detaches from the solid surface and moves up along the surface of the droplet. The compressed liquid is now free to spread or "jet" out laterally and relieve the contact pressures (Fig. 4c). In experiments, however, actual jetting is not observed until the contact angle ϕ has reached a value significantly greater than the theoretical critical value ϕ_c; the reasons for this behavior are not yet fully understood. The maximum lateral jetting velocity is many times greater than the impact velocity, but theoretical prediction of its value is still uncertain.

Figure 5 shows high-speed photographs at various stages of the impact between a solid projectile and a gelatine droplet in a laboratory apparatus. In this case the impact velocity was 110 m/s (360 ft/s) and the maximum jetting velocity 1170 m/s (3840 ft/s).

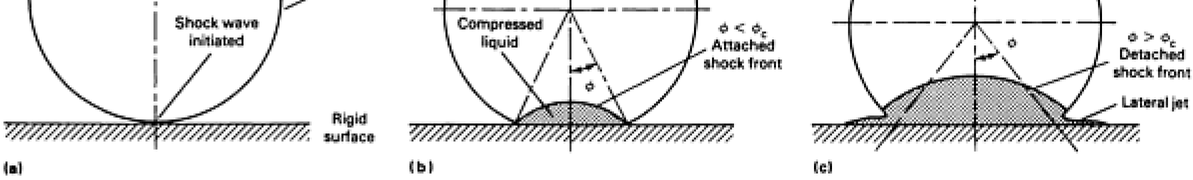


Fig. 4 Idealized diagram of the early stages of liquid drop impact. (a) Initial contact. (b) Compressible stage with attached shock front. (c) Detached shock and jetting stage. See text for detailed discussion of these three stages.

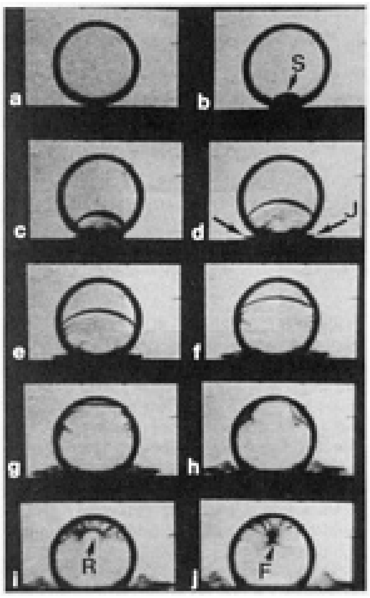


Fig. 5 High-speed photographs of the impact between a 10 mm (0.4 in.) diam two-dimensional droplet of gelatin and a metal slider moving at 110 m/s (360 ft/s) at intervals of 1 μs. S denotes the shock front, first seen in frame b. Jetting is just visible in frame c and is labeled J in frame d. Note the reflection of the shock as a rarefaction wave R in frames h through j, causing a region of cavitation indicated by F. In these photos, there has not yet been time for any gross spreading out of the droplet to occur. Courtesy of J.E. Field, Cavendish Laboratory, University of Cambridge, United Kingdom

The liquid/solid interaction is further complicated when the solid surface becomes deformed from erosion, usually exhibiting jagged peaks and craters. Then both the pressures and jetting patterns will be affected by where the initial contact takes place, and by the size of the droplet in relation to topographic features. For example, a drop falling on a peak or slope may not develop full impact pressure; one falling in a crater may produce increased pressures due to shock wave collisions or "shaped charge" effects. This may explain the characteristic pitted appearance because if even shallow pits are formed, subsequent material loss occurs preferentially in the pits and continues to deepen them.

Material Response--Development of Damage. As described above, the solid surface is subjected to a multitude of sharp pressure pulses and jetting outflows, each of very short time duration and acting on a very small area. What then happens to the solid material is hard to generalize because it will depend on whether the solid is ductile or brittle, on its microstructure, and on whether the impacts are severe enough to produce single-impact damage. Adler (Ref 3) lists the primary causes of damage as direct deformation, stress wave propagation, lateral outflow jetting, and hydraulic penetration. At impact, the formation of the shock front in the liquid is accompanied by corresponding stress waves propagating into the solid; the solid response is therefore also impulsive and governed by its dynamic rather than static mechanical properties.

In ductile materials, a single intense impact may produce a central depression, with a ring of plastic deformation around it where the jetting outflow may remove material by a tearing action (Fig. 6). With less intense but repeated impacts, there is no immediate material loss, but randomly disposed dimples gradually develop, and the surface undergoes gradual deformation (often characterized by twinning) and work hardening. Metallographic and x-ray diffraction studies have shown that during this "incubation stage," these effects eventually extend to 30 to 50 μm below the surface, thereafter remaining about the same as actual erosion (material loss) then begins and progresses. The material loss may occur through propagation of fatiguelike cracks that eventually intersect to release erosion fragments. The fractures have often been described as transgranular. This process can be assisted by tearing that is due to increased liquid forces on irregular surface steps and fissures. In materials with pronounced nonuniform structure, damage will initiate at weak spots or in the weaker components. Figures 7 and 8 illustrate the character of erosion damage in a martensitic stainless steel (AISI type 403) and Stellite 6B, respectively.

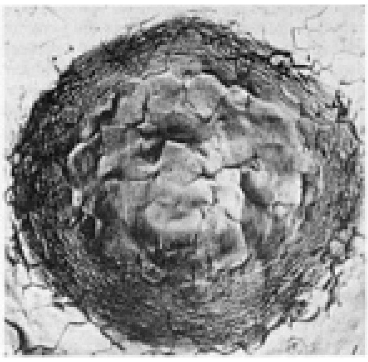


Fig. 6 Deformation due to a single impact on aluminum impacted by a short discrete jet of water at 750 m/s (2460 ft/s). Note the central depression, which is of similar diameter to the impacting jet, and the circumferential surface ripples surrounding it. Courtesy of J.E. Field, Cavendish Laboratory, University of Cambridge, United Kingdom

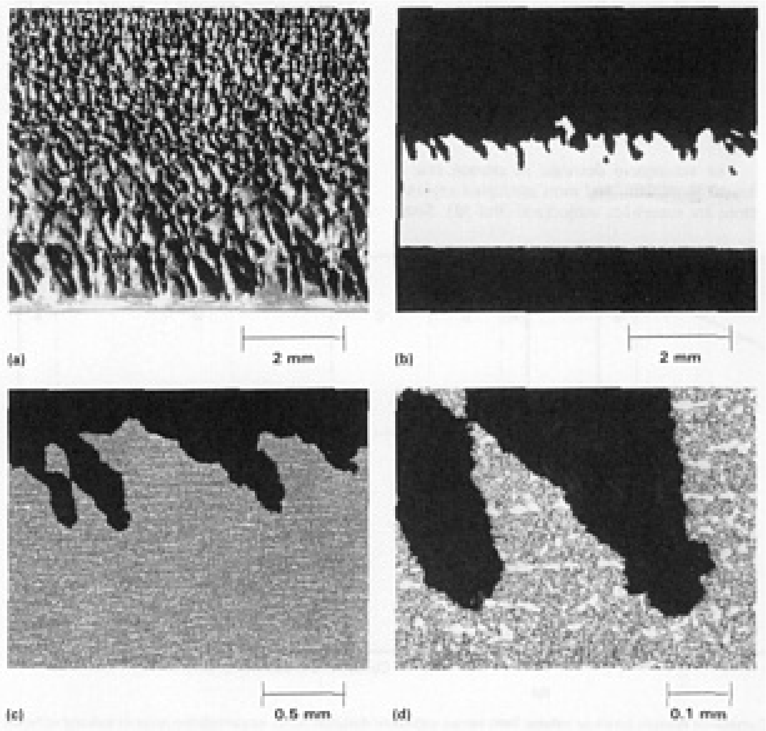


Fig. 7 Character of erosion in type 403 martensitic stainless steel. (a) Macrograph of eroded area. 10×. (b) Unetched section. 10×. (c) Section through several pits. GRARD II etch. 50×. (d) Enlarged portion of (c). 200×. Courtesy of Westinghouse Electric Corporation

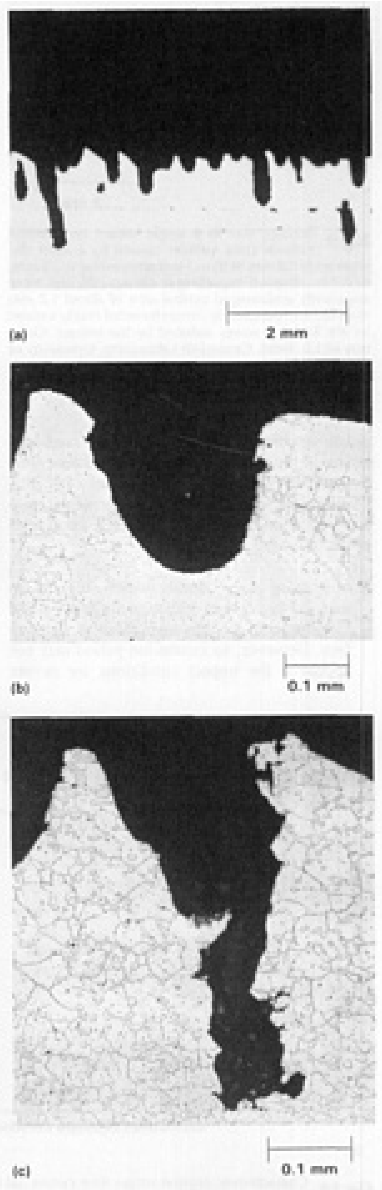


Fig. 8 Character of erosion in Stellite 6B alloy. (a) Unetched section. 10×. (b) and (c) Etched sections through specific pits. Both at 200×

In brittle materials, circumferential cracks may form around the impact site that are caused by tensile stress waves propagating outward along the surface (Fig. 9). In thin sheets subjected to impacts, material can spall off the inside surface due to the compressive stress wave from the impact reflecting there as a tensile wave.

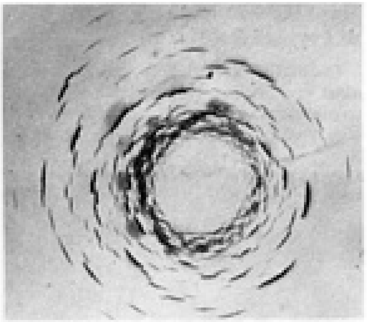


Fig. 9 Damage due to a single impact on a brittle material (zinc sulfide) caused by a short discrete jet of 0.8 mm (0.03 in.) (corresponding to a 5 mm, or 0.2 in., droplet) impacting at 300 m/s (985 ft/s). Note apparently undamaged central area of about 1.2 mm (0.05 in.) surrounded by circumferential cracks caused by the Rayleigh waves induced by the impact. Courtesy of J.E. Field, Cavendish Laboratory, University of Cambridge, United Kingdom, from Ph.D. thesis by D. Townsend

For materials and composite structures used in aerospace applications, it is difficult to generalize damage mechanisms. For example, in thermosetting polymers or chopped fiber-reinforced composites, the damage takes the form of chunking (removal of large-size lumps of material) from the surface (Ref 25). Furthermore, the initial measure of damage for radome and infrared window materials is not gross material removal, but impairment of electromagnetic transmission characteristics, which leads to loss of function. In the extreme, however, damage can be catastrophic, as shown in Fig. 2.

Corrosion Interactions. More research on corrosion interactions has been done for cavitation than for liquid impingement, but the general observations described below can apply to both.

In the early days before high impact pressures were understood, it was often supposed that liquid impingement as well as cavitation damage had to be largely or significantly corrosive in nature. It is now recognized, and has been proven experimentally, that such erosion can occur without any corrosive component. Moreover, under impingement of high intensity, material loss can occur so rapidly that corrosion--even if otherwise possible--does not have time to play a role. Nevertheless, at intermediate mechanical intensities there is opportunity for corrosion to weaken the material and facilitate its removal by the mechanical impact forces. Several investigators have shown some parallels between erosion and corrosion fatigue behavior.

Time Dependence of Erosion Rate

Qualitative Description. Liquid impingement erosion shares with cavitation erosion a unique characteristic that makes accurate long-term predictions or extrapolations of erosion nearly impossible. This is the very nonlinear nature of the progress of erosion with time, under constant impingement conditions (Ref 38). Qualitatively, the so-called "erosion-time pattern" depicted in Fig. 10 is generally composed of the following stages:

- *Incubation stage,* during which little or no material loss occurs, although roughening and metallurgical changes take place in the surface. However, an incubation period may not appear if the impact conditions are severe enough for each single impact to cause material loss
- *Acceleration stage,* during which the erosion rate increases rapidly to a maximum
- *Maximum rate stages,* during which the erosion rate remains constant or nearly so. The erosion rate for this stage is most commonly quoted as a single-number result of an erosion test. However, some tests show only a fleeting peak in the erosion rate-versus-time curve, with no prolonged steady rate
- *Deceleration (or attenuation) stage,* during which the erosion rate declines to some fraction (often from $\frac{1}{4}$ to $\frac{1}{2}$) of the maximum rate
- *Terminal (or final steady-state) stage,* during which the rate remains constant once again indefinitely. However, some tests do not show this stage, and the erosion rate either continues to decline or goes into a series of fluctuations. With some brittle materials or coatings, the rate can increase in what is called a

"catastrophic stage"

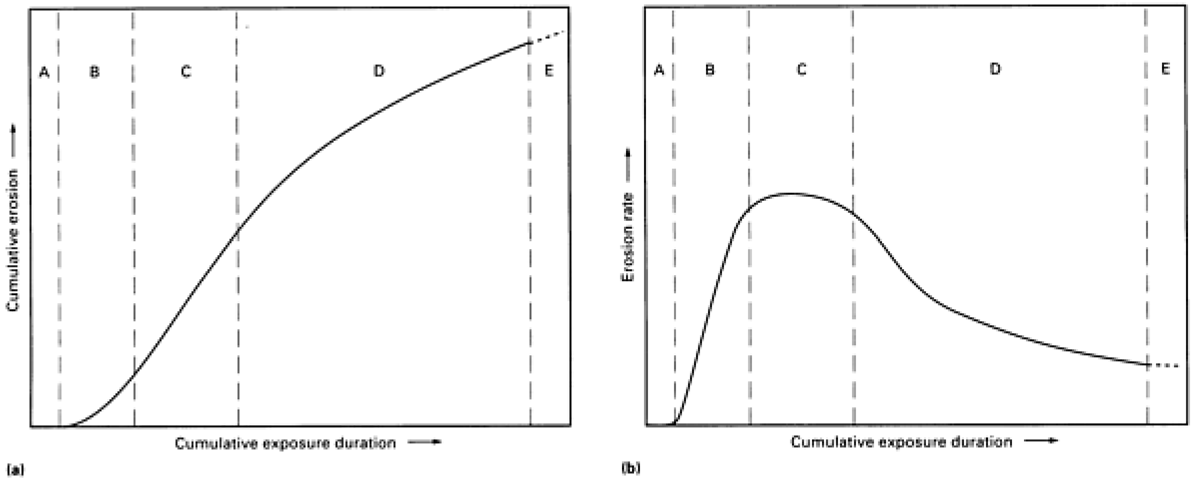


Fig. 10 Characteristic erosion versus time curves. (a) Cumulative erosion (mass or volume loss) versus exposure duration (time, or cumulative mass or volume of liquid impinged). (b) Corresponding instantaneous erosion rate versus exposure duration obtained by differentiating curve (a). The following stages have been identified thereon: A, incubation stage; B, acceleration stage; C, maximum rate stage (sometimes called first steady-state stage); D, deceleration stage; and E, terminal or final steady-state stage, if assumed to exist.

Figure 11 shows some nontypical curves that are sometimes encountered. All of the curves in Fig. 10 and 11 have been derived from actual test curves in the literature.

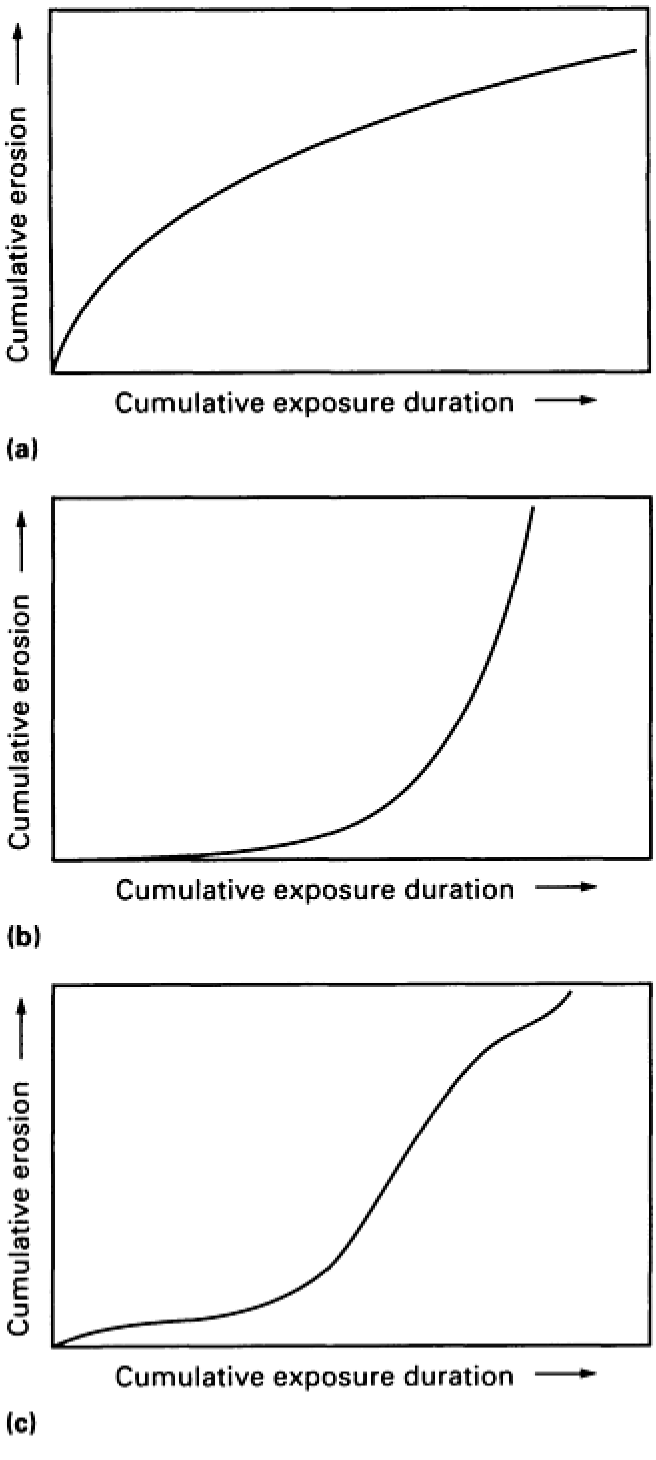


Fig. 11 Less common types of cumulative erosion versus time curves sometimes obtained. (a) Curve without incubation or acceleration stages, with continuously decreasing rate (obtained in this case with very small droplets and very high impact velocity). (b) Curve with continuously increasing erosion rate, resulting in catastrophic damage (obtained in this case on titanium carbide). (c) Curve with fluctuations in erosion rate (obtained in this case on a titanium alloy)

Reasons for Time Dependence. The incubation and acceleration stages are easy to explain in qualitative terms if one postulates that the removal of erosion fragments results from a fatiguelike failure mechanism. Then many impacts must occur in one area for a fragment to be loosened from the surface. The statistical nature of the process then results in a gradual transition--the acceleration period--from the incubation stage to the steady-state (maximum rate) stage.

The subsequent decrease in erosion rate is harder to explain, and most attempted explanations are somewhat conjectural (Ref 38). Some have also been based on the statistics of the damage mechanisms, combined with changes in the surface properties brought on by erosion itself. Some are based on the topographical changes in the surface: as the surface is roughened, the surface area is increased, and more energy is needed to continue erosion. Also, liquid drops will now tend to impact on the peaks or the slopes of the roughened surface; in both cases the impact pressures may be reduced. Finally, the liquid retained in erosion craters has been supposed to cushion and protect the surface.

Implications for Testing and Prediction. Clearly, this complicated time dependence of erosion, and the variations thereof, make it very difficult to define a single meaningful test result of an erosion test, or assign from that an erosion resistance of a material, or predict long-term erosion behavior from a short test. Many authors in the literature have proposed mathematical formulations for the erosion-time curve (or portions of it). Some of these have been essentially empirical, some have been based on proposed analytical models for the erosion process; however none so far has achieved general acceptance. See the discussion of "Comprehensive Prediction Methods and Erosion Theories" later in this article for additional information.

For the purpose of reporting test results, ASTM Standard Practice G 73 for Liquid Impingement Erosion Testing (Ref 39), as well as ASTM Standard Method G 32 for Vibratory Cavitation Erosion Testing (Ref 40) recommend that curves of cumulative mass loss versus time be shown in a test report, since any other parameters (for example, erosion rates) must be derived from that. For tabular data and comparisons (between materials or test conditions), ASTM G 73 prescribes tabulating the maximum erosion rate and a nominal incubation period, which is simply the intercept on the time axis of the maximum erosion rate line. In order to describe the longer-term behavior, some authors also use the terminal (final steady-state) erosion rate and some parameter that defines the intersection of that line with the maximum-rate line or the erosion axis (Fig. 12). Note that even this simplified model of the erosion-time pattern, consisting of three straight-line segments, requires four experimentally or analytically derived parameters to define it.

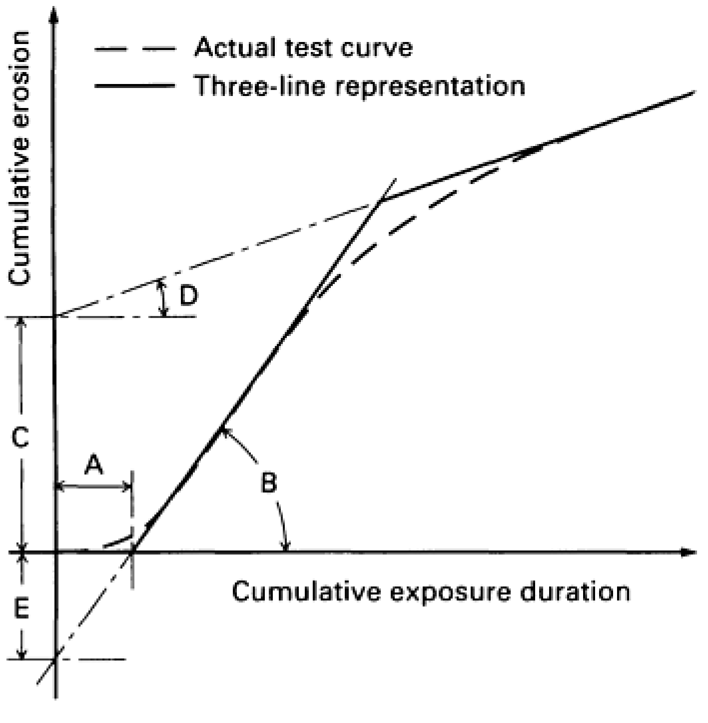


Fig. 12 The erosion versus time curve from Fig. 10, showing some numerical parameters that may be recorded to characterize the test results. A, nominal incubation period; B, slope representing maximum erosion rate; C, *y*-axis intercept of terminal erosion rate line; D, slope representing terminal erosion rate. ASTM G 73 (Ref 39) specifies that at least A and B be tabulated in any test report. Some authors have used E (*y*-axis intercept of maximum erosion rate line) in place of A; it tends to be constant when some parameters (for example, impact velocity) are varied.

The author's experience, in attempting to correlate test data from various sources, has been that the maximum erosion rate is the most predictable by empirical relationships. However, this parameter alone cannot be used to predict long-term

erosion in the absolute sense, and there are indications that it does not predict it well even in the relative sense. Thus, more work is certainly needed in reaching a generalized method for long-term erosion prediction.

Factors Affecting Erosion Severity

In order to understand the mechanics of erosion and to subsequently predict the erosion behavior of various materials, there are a number of critical parameters that must be examined. Following a brief discussion of the dimensionless parameters used for characterizing erosion, this section will review critical empirical observations regarding both impingement variables (velocity, impact angle, droplet size, and physical properties of liquids) and erosion resistance of materials, including the correlation between erosion resistance and mechanical properties and the effects of alloying elements and microstructure. Empirical erosion prediction equations developed from ASTM-sponsored test programs are also described.

Dimensionless Parameters for Describing Erosion

When a physical phenomenon is fully understood, it should be possible to describe it mathematically in terms of a functional relationship between a set of dimensionless parameters. This ensures dimensional consistency in the functional relationships, makes the equations independent of the set of units of measurement adopted, and can help ensure--by use of dimensional analysis--that the correct number of independent parameters are included for each dependent parameter of interest. Unfortunately, no such complete relationship for predicting erosion has been generally accepted to date, although some authors have made attempts toward it. Although an extended discussion of this topic is beyond the scope of this article, two dependent variables of interest in dimensionless form will be introduced. These are:

- The "rationalized erosion rate," R_e, which is defined as the volume of the target material lost divided by the volume of liquid impinged (where both pertain to the same area and the same time interval)
- The "rationalized incubation period," N_0, which is defined as the number of stress pulses experienced by a typical point on the target surface during the incubation period

In the following discussions, the "erosion rate" referred to is always the "maximum erosion rate" (see the previous section of this article on "Time Dependence of Erosion Rate"). The empirical data regarding incubation periods are much less abundant, but several studies have suggested that the nominal incubation period defined in Fig. 12 is approximately inversely proportional to the maximum erosion rate and, moreover, the product of the rationalized quantities R_eN_0 is on the order of unity (Ref 41).

Impingement Variables

Velocity Dependence/Threshold Considerations. It is conceptually very attractive to suppose that there is a "threshold velocity"--dependent on the material--below which no erosion would occur, analogous to the endurance limit in fatigue. Several investigators have presented evidence for such a threshold (Ref 22, 42), which may also depend on droplet size (Ref 43). However, the author has found that most test data, from many sources, seem to better fit a simple power law in which the rationalized maximum erosion rate varies with about the 4th to 5th power of impact velocity (Ref 44). For brittle materials, exponents as high as 6 to 9 have been reported (Ref 45). At low impact velocities, the incubation period may become so long that no actual material loss takes place in a reasonable testing or operating time, giving the appearance of a threshold. Thus, the question of a threshold phenomenon is not yet firmly settled.

Dependence on Impact Angle. It is generally considered that, to a first approximation, erosion depends on only the normal component of the impact velocity; thus, because of the strong dependence on impact velocity, erosion is reduced strongly as impacts become more glancing. Some investigators have suggested a small added contribution from the tangential component, and intuitively one might suspect that once a surface becomes roughened by erosion, the effect of the tangential component would become more pronounced.

Dependence on Droplet Size. By and large, test data show that R_e decreases with drop size (Ref 41, 43); that is, a given total amount of liquid does less damage if divided into smaller drops, even though this implies a greater number of impacts on the surface. There is no obvious explanation for this phenomenon. One suggested explanation is that it is due to the shorter time duration of each pressure pulse with smaller drops; another postulates that it is a material-related size

effect similar to those in fatigue notch sensitivity or in metal cutting, where the spatial extent of imposed stresses must exceed some characteristic dimension.

Dependence on Liquid Properties. Most liquid impact erosion tests have been conducted with water at near normal atmospheric conditions. Knowledge of how erosion varies with the physical properties of the liquid is necessary if one wants to extend empirical relationships to other liquids and conditions, and also if one wants to construct a true analytical model that is dimensionally consistent. The results of one investigation (Ref 46) suggested dependencies on approximately the 2nd to 2.5th power of liquid density and the $\frac{1}{2}$ to $\frac{3}{4}$ power of the inverse of viscosity; the dependence on acoustic velocity (which theoretically enters into impact pressure as well as acoustic impedance relationships) remained unclear. More work in this area is desirable.

Erosion Resistance

What Is Erosion Resistance? In the absence of any widely accepted, dimensionally consistent, analytical model for the erosion process, it is impossible to specify an independent definition, dimensions, or units for the property of erosion resistance. Although a concept of "erosion strength" having the units of a mechanical strength or strain energy property has been proposed (Ref 42), this has not gained wide acceptance.

All one can do is to conduct erosion tests and compare the results between different materials. Generally one material is selected as a reference material and results for the others are referred to that. ASTM G 40 (Ref 1) defines a "normalized erosion resistance" as follows: "The volume loss rate of a test material, divided into the volume loss rate of a specified reference material similarly tested and similarly analyzed . . ." In Ref 44 the reference material is austenitic stainless steel of hardness HV 170. ASTM G 73 (Ref 39) defines an "erosion resistance number" based on up to three reference materials covering a range of resistances. Some recent cavitation test programs have adopted stainless steel type 308 weld overlay as the reference.

Correlations with Mechanical Properties. Hardness has a strong effect on erosion resistance, but by itself is not a sufficient indicator. For the same or similar materials, erosion resistance generally varies with about the 2nd to 2.5th power of Vickers hardness number, but that does not necessarily apply when different types of materials are compared. For example, austenitic stainless steels have higher resistance than martensitic stainless steels of the same hardness, and cobalt-base alloys (for example, Stellites) have an even higher resistance in relation to hardness (Ref 44).

Other parameters based on mechanical properties have been proposed as indicators of erosion resistance. These include the strain energy to fracture parameter, and the so-called "ultimate resilience" ($s_u^2/2E$), where s_u is ultimate strength and E is modulus of elasticity; neither is very successful over a broad range of materials (Ref 44).

Since liquid impingement and cavitation erosion is due to repeated stress pulses and thus related to fatigue, some authors have tried unsuccessfully to correlate resistance with endurance limit or finite fatigue life. Recently, however, Richman and McNaughton (Ref 47) demonstrated good correlation with cyclic deformation properties for a number of alloys and pure metals.

Several authors (Ref 48, 49, 50) who have attempted theoretical analyses of the erosion process have come up with much more complicated parameters to define the resistance of a material, but these parameters tend to be difficult to evaluate or, in some cases, fail to predict the empirically observed dependencies.

Effects of Alloying Elements and Microstructure. The coverage in this section is based on cavitation erosion as well as liquid impingement studies. See also the article "Cavitation Erosion" in this Volume for additional information.

Improved erosion resistance has been associated with alloying elements such as chromium, manganese, and cobalt. The effect of nickel is inconsistent. For example, a steel containing 12% Cr and 4% Ni is better than straight 12% Cr steel, but some studies have found nickel deleterious. Fine microstructure is advantageous and so is the ability of the surface layer to work harden as a result of impact-induced deformation. The extremely high erosion resistance of Stellite (cobalt-chromium-tungsten alloy) has been variously explained by a microstructure consisting of small hard carbide particles in a strong but more ductile matrix, or by crystallographic transformations induced by the liquid impacts.

Very high resistance has been reported for chromium-manganese steels (about 10% Cr and 12% Mn) that undergo austenitic-martensitic phase transformation under impingement (Ref 51). Akhtar *et al.* (Ref 52) investigated stainless steel

as well as nickel and cobalt-base weld overlays and found certain optimum composition parameters for each, but not the same for all. They also found that stress-induced martensitic transformations control erosion of austenitic stainless steels.

Simoneau *et al.* (Ref 53) found that low stacking-fault energy is the key to high erosion resistance in austenitic stainless steels as well as cobalt-base alloys. They have developed a new class of cobalt-containing austenitic stainless steels with erosion resistance comparable to cobalt-base alloys such as Stellite. The low stacking-fault energy results in a "cavitation-induced hardened surface layer" featuring "planar deformation, fine twinning, high strain hardening and high fatigue resistance, smoother eroded facies, fine eroded particle size."

Comprehensive Prediction Methods and Erosion Theories

Reference 44 gives a simple empirical prediction equation for the maximum erosion rate in liquid impingement erosion, as well as a simplified time dependence factor based on the assumption that the rate continues to decline indefinitely from its maximum. Slightly more elaborate equations for maximum erosion rate and for incubation period are given in Ref 39 and 41, based on an interlaboratory test program sponsored by ASTM Technical Committee G-2 on Wear and Erosion. These are:

$$\log R_e = 4.8 \log V - \log NER - 16.65 + 0.67 \log d + 0.57\ J - 0.22\ K$$

$$\log N_0 = -4.9 \log V + \log NOR + 16.40 - 0.40\ J$$

where R_e and N_0 are the rationalized erosion rate and incubation period, respectively, V is impact velocity (m/s) normal to the target surface, d is drop diameter (mm), J is 0 for droplet impingement and 1 for repetitive impacts against cylindrical jets, K is 0 for flat targets and 1 for round or curved targets, and NER and NOR are the "erosion resistance number" and "incubation resistance number," respectively, as defined in Ref 39. The latter are normalized resistance values with respect to AISI type 316 stainless steel of hardness about HV 165.

The purely empirical formulations given above are based on a wide collection of test data from different laboratories with widely different impingement conditions. With all the variabilities that occur, no erosion prediction can be very accurate, but the above equations have been shown to predict the maximum erosion rate within a factor of less than 3 for most test results with essentially uniform drop size and impact velocity conditions. The error for incubation period is greater. In real situations with a distribution of drop sizes and impact velocities, prediction is still more unsure.

These equations still do not suffice for long-term prediction. Although Ref 44 offers a very tentative correction factor for the time dependence, it is based on the assumption of a continually decreasing erosion rate. A more conservative approach may be to assume that the "terminal erosion rate" is approximately 30 to 60% of the maximum rate and continues indefinitely. Rao and Buckley (Ref 54) have also tried to unify the time dependence from many test data.

Some other comprehensive prediction methods and theories should be mentioned. Springer (Ref 48) presents semi-analytically based relationships, but they overestimate incubation periods and underestimate erosion rates by about 4 orders of magnitude; this was subsequently attributed by the author to a numerical error in calculating one of the empirical constants from experimental data (Ref 55).

Ruml (Ref 56) has proposed an empirical equation for the whole erosion-time relationship in terms of dimensionless parameters; many empirical constants are required. Karimi and Leo (Ref 49) offer a new analytical model for erosion rates and their time dependence, but it involves statistical parameters difficult to determine. Other recent contributions toward representation of long-term erosion include those of Beckmann (Ref 50), Shubenko and Kovalsky (Ref 57), and Stanіša *et al.* (Ref 58). Perhaps these new efforts will eventually result in a practical generalized prediction scheme.

Test Methods for Erosion Studies

Erosion tests are conducted to evaluate materials as well as to study the effects of other variables on erosion. Basic studies of single impacts have been conducted with "liquid gun" devices, in which a small quantity of liquid is ejected through a carefully designed nozzle and impacts a stationary specimen (Ref 33), or by projecting a solid target against a stationary liquid or gelatine body (Ref 37).

However, for multiple-impact studies and for evaluating the resistance of materials, the usual approach is to attach the specimen(s) to the periphery of a rotating disc or arm, such that in their circular path they repeatedly pass through and impact against liquid jets, sprays, or simulated rain drops (Ref 20, 21, 39, 41, 42). The velocity of the specimen then determines the impact velocity. Such test facilities range from small laboratory apparatus with specimen velocities of up to about 200 m/s (655 ft/s) to large self-contained facilities (some with vacuum-controlled test chambers); some of the latter are capable of impact velocities up to 1000 m/s (Mach 3, or 3300 ft/s). ASTM G 73 (Ref 39) gives comprehensive guidelines for conducting this type of test and for analyzing the data.

The successful selection and development of improved materials and coatings for rain erosion have been largely based on rotating arm tests. Service experience and in-flight tests have confirmed that the laboratory tests predict the correct relative erosion resistance as well as the failure modes. For high supersonic rain erosion studies, however, specimens have been attached to rocket sleds propelled at speeds up to 1700 m/s (5580 ft/s) through an artificial rain field (Ref 45). Wind tunnels have also been adapted for rain erosion tests.

As noted earlier, cavitation erosion test methods such as the vibratory method (Ref 40), the cavitating jet method, and the submerged rotating disk method have also been used to screen materials for service under liquid impingement conditions.

Means for Combatting Erosion

Modification of Impingement Conditions. If possible, the geometry and/or fluid dynamics should be modified to reduce the amount of liquid impacting the exposed surfaces, to reduce the impact velocity of the droplets (or change the impact angle to reduce the normal component of impact velocity), to reduce the droplet size, or to reduce time of operation under the most severe conditions. Efforts in all of these directions are made in steam turbines as well as in aircraft and missiles, but details are beyond the scope of this discussion.

Material Selection and Protective Shielding. Some of the material aspects have already been covered in earlier sections of this article. Figure 13 provides an overview of relative erosion resistance of metallic alloys, based primarily on a series of results (from both impingement and cavitation tests) examined in Ref 44. Some items from recent investigations (Ref 52, 53, 59, 60) have been added; most of these were cavitation tests. Because of the inconsistencies inherent in erosion testing, this can be used only as a rough guide. Some investigators have found that the relative resistances of some materials depended on the impact velocity, implying significantly different velocity dependencies for them.

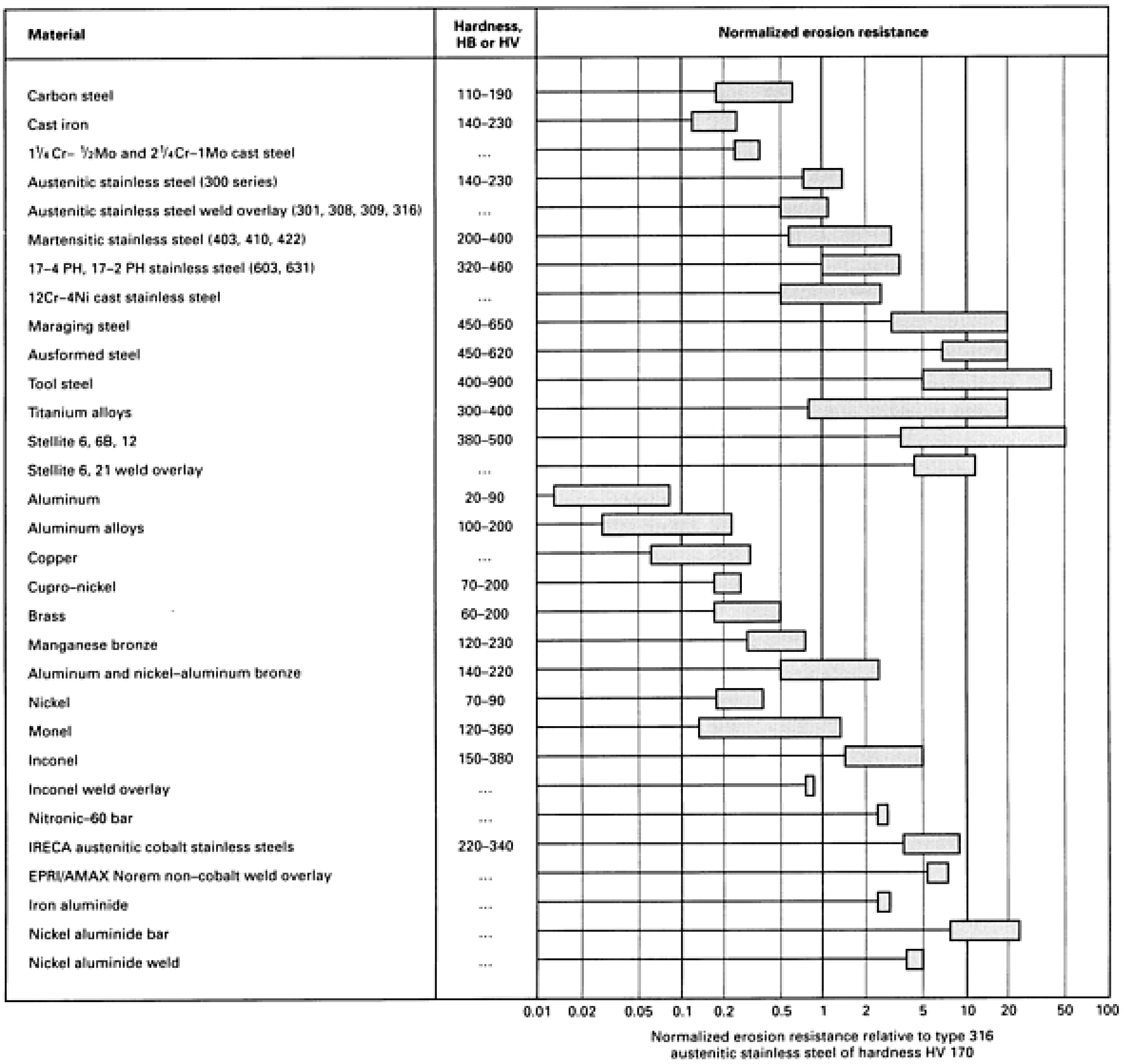


Fig. 13 Normalized erosion resistance of various metals and alloys relative to type 316 austenitic stainless steel of hardness 170 HV (the erosion resistance number according to ASTM G 73). Data deduced from many sources in the literature including both impingement and cavitation tests. It must be cautioned that erosion test data are not very consistent, and the information herein should be used only as a rough guide.

Shielding or cladding approaches for erosion protection have usually employed a harder layer than the base material, or one that by virtue of its composition or microstructure is more resistant to erosion (and corrosion). However, another approach that can be successful in low-intensity environments is the use of elastomeric coatings, which, by virtue of their elasticity, reduce the magnitude of the impact pressures; this approach is widely used against rain erosion at moderate speeds.

Table 1 lists some of the materials and coatings used for rain erosion applications, in order of preference as specified by Schmitt (Ref 25).

Table 1 Materials in use for rain erosion protection

"Recommended" as well as "not recommended" materials are listed in order of preference or resistance. "Not recommended" materials may have good erosion resistance but other undesirable properties such as poor shock resistance.

Materials for aircraft radomes	
Recommended	**Not recommended**
Elastomeric coatings:	
Polyurethane	Silicone
Fluorocarbon	Acrylic
Neoprene	Polyester
Bulk plastics:	
Modified polyphenylene oxide	Tetrafluoroethylene
Polycarbonate	Polyimide
Polyethylene	Epoxy
Polyacetal	Acrylonitrile-butadiene-styrene (ABS)
Nylon	
Modified acrylic	
Reinforced composites:	
Glass cloth reinforced epoxy	Chopped glass reinforced epoxy or polyester
Filament wound glass epoxy	Kevlar-epoxy
Quartz cloth reinforced epoxy	

Materials for missile radomes	
Recommended	**Not recommended**
Monolithic ceramics:	
Hot-pressed silicon nitride	Slipcast fused silica
Alumina	Boron nitride
Beryllia	
Cordierite	
Reaction-sintered silicon nitride	
Reinforced ceramics:	
1-D celcon polyacetal silica	. . .
Alumina cloth-silica matrix	
3-D reinforced silica-silica	

Materials for optical and infrared domes	
Recommended	**Not recommended**
Glasses and ceramics:	
Sapphire	. . .
Spinel	
Calcium aluminosilicate	
Calcium borosilicate	
Magnesium fluoride	
Clear fused silica	
Chalcogenides:	
Germanium	. . .
Zinc sulfide	
Gallium arsenide	
Zinc selenide	

Source: Ref 25

References

1. "Standard Terminoloy Relating to Wear and Erosion," G40, *Annual Book of ASTM Standards,* ASTM
2. C.M. Preece, Ed., *Treatise on Materials Science and Technology,* Vol 16, *Erosion,* Academic Press, 1979
3. W.F. Adler, The Mechanics of Liquid Impact, *Treatise on Materials Science and Technology,* Vol 16, *Erosion,* C.M. Preece, Ed., Academic Press, 1979, p 127-183
4. J.H. Brunton and M.C. Rochester, Erosion of Solid Surfaces by the Impact of Liquid Drops, *Treatise on Materials Science and Technology,* Vol 16, *Erosion,* C.M. Preece, Ed., Academic Press, 1979, p 185-248
5. F.P. Bowden, Organizer, Deformation of Solids by the Impact of Liquids (and Its Relation to Rain Damage in Aircraft and Missiles, Blade Erosion in Stream Turbines, Cavitation Erosion), *Philos. Trans. R. Soc. A,*

Vol 260 (No. 1110), 1966

6. *Symposium on Erosion and Cavitation,* STP 307, ASTM, 1962
7. *Erosion by Cavitation or Impingement,* STP 408, ASTM, 1967
8. *Characterization and Determination of Erosion Resistance,* STP 474, ASTM, 1970
9. *Erosion, Wear, and Interfaces with Corrosion,* STP 567, ASTM, 1974
10. W.F. Adler, Ed., *Erosion: Prevention and Useful Applications,* STP 664, ASTM, 1979
11. A.A. Fyall and R.B. King, Ed., *Proceedings of the Rain Erosion Conference* (Meersburg, West Germany), 5-7 May 1965, Royal Aircraft Establishment, Farnborough, England
12. A.A. Fyall and R.B. King, Ed., *Proceedings of the Second Meersburg Conference on Rain Erosion and Allied Phenomena,* 16-18 Aug 1967, Royal Aircraft Establishment, England
13. A.A. Fyall and R.B. King, Ed., *Proceedings of the Third International Conference on Rain Erosion and Allied Phenomena,* Royal Aircraft Establishment, England, 1970
14. A.A. Fyall and R.B. King, Ed., *Proceedings of the Fourth International Conference on Rain Erosion and Allied Phenomena,* Royal Aircraft Establishment, England, 1974
15. *Proceedings of the Fifth International Conference on Erosion by Liquid and Solid Impact* (ELSI-V), Cavendish Laboratory, University of Cambridge, England, 1979
16. J.E. Field and N.S. Corney, Ed., *Proceedings of the Sixth International Conference on Erosion by Liquid and Solid Impact* (ELSI-VI), Cavendish Laboratory, University of Cambridge, England, 1983
17. J.E. Field and J.P. Dear, Ed., *Proceedings of the Seventh International Conference on Erosion by Liquid and Solid Impact* (ELSI-VII), Cavendish Laboratory, University of Cambridge, England, 1987
18. G.C. Gardner, Events Leading to Erosion in the Steam Turbine, *Proc. Inst. Mech. Eng.,* Vol 178, Part 1 (No. 23), 1963-1964, p 593-623
19. M.J. Moore and P. Schulpher, Conditions Producing Concentrated Erosion in Large Steam Turbines, *Wet Steam-3, Proc. Inst. Mech. Eng.,* Vol 184, Part 3G (No. III), 1969-1970, p 142-158
20. D.E. Elliott, J.B. Marriott, and A. Smith, Comparison of Erosion Resistance of Standard Steam Turbine Blade and Shield Materials on Four Test Rigs, STP 474, ASTM, 1970, p 127-161
21. D.W.C. Baker *et al.,* The Erosion Resistance of Steam Turbine Blade and Shield Materials, *Proceedings of the Second Meersburg Conference on Rain Erosion and Allied Phenomena,* 16-18 Aug 1967, Royal Aircraft Establishment, England, p 449-515
22. D. Pollard, M.J. Lord, and E.C. Stockton, An Evaluation of Low Pressure Steam Turbine Blade Erosion, *Proceedings of the Design Conference on Steam Turbines for the 1980s,* Institution of Mechanical Engineers, 1979, p 413-419 [also *GEC J. Sci. Technol.,* Vol 49 (No. 1), 1983, p 29-34]
23. J.A. Krzyżanowski, Experience in Predicting Steam Turbine Blade Erosion, *Proceedings of the Seventh International Conference on Erosion by Liquid and Solid Impact* (ELSI-VII), Cavendish Laboratory, University of Cambridge, England, 1987, p 11-1 to 11-7
24. A.A. Fyall, Rain Erosion--A Special Radome Problem, Chapt. 8, *Radome Engineering Handbook,* Marcel Dekker, 1970, p 461-570
25. G.F. Schmitt, Jr., Air Force Materials Laboratory, personal communication, 1991
26. D.A. Summers, Practical Applications of Erosion Processes, *Treatise on Materials Science and Technology,* Vol 16, *Erosion,* C.M. Preece, Ed., Academic Press, 1979, p 395-441
27. D.A. Summers, *Jetting Technology,* Elsevier, 1991
28. J. Ducreux, The Influence of Flow Velocity on the Corrosion-Erosion of Carbon Steel in Pressurized Water, *Water Chemistry 3,* British Nuclear Engineering Society, London, 1983, p 227-233
29. R.G. Keck and P. Griffith, Models and Equations for the Prediction of Erosive-Corrosive Wear in Steam Extraction Piping, Paper 87-JPGC-Pwr-35, American Society of Mechanical Engineers, 1987
30. P.A. Coulon, Erosion-Corrosion in Steam Turbines Part II: A Problem Largely Resolved, *Lubr. Eng.,* Vol 42 (No. 6), 1986, p 357-362
31. N. Henzel, W. Kastner, and B. Stellwag, Erosion Corrosion in Power Plants under Single- and Two-Phase

Flow Conditions--Updated Experience and Proven Counteractions, *Proceedings of the American Power Conference,* 1988, p 992-1000

32. H. Keller, Erosionskorrosion an Nassdampf-turbinen, *VGB Kraftwerkstechnik,* Vol 54 (No. 5), 1974, p 292-295
33. J.H. Brunton, Deformation of Solids by Impact of Liquids at High Speeds, STP 307, ASTM, 1962, p 83-98
34. F.J. Heymann, High-Speed Impact between a Liquid Drop and a Solid Surface, *J. Appl. Phys.,* Vol 40 (No. 13), 1969, p 5113-5122
35. J.E. Field, M.B. Lesser, and P.N.H. Davies, Theoretical and Experimental Studies of Two-Dimensional Liquid Impact, *Proceedings of the Fifth International Conference on Erosion by Liquid and Solid Impact* (ELSI-V), Cavendish Laboratory, University of Cambridge, England, 1979, p 2-1 to 2-8
36. P.H. Pidsley, A Numerical Investigation of Water Drop Impact, *Proceedings of the Sixth International Conference on Erosion by Liquid and Solid Impact* (ELSI-VI), Cavendish Laboratory, University of Cambridge, England, 1983, p 18-1 to 18-6
37. J.E. Field, J.P. Dear, and J.E. Ogren, The Effects of Target Compliance on Liquid Drop Impact, *J. Appl. Phys.,* Vol 65 (No. 2), 15 Jan 1989, p 533-540
38. F.J. Heymann, On the Time Dependence of the Rate of Erosion due to Impingement or Cavitation, STP 408, ASTM, 1967, p 70-110
39. "Standard Practice for Liquid Impingement Erosion Testing," G 73, *Annual Book of ASTM Standards,* ASTM
40. "Standard Method of Vibratory Cavitation Erosion Test," G 32, *Annual Book of ASTM Standards,* ASTM
41. F.J. Heymann, Conclusions from the ASTM Interlaboratory Test Program with Liquid Impact Erosion Facilities, *Proceedings of the Fifth International Conference on Erosion by Liquid and Solid Impact* (ELSI-V), Cavendish Laboratory, University of Cambridge, England, 1979, p 20-1 to 20-10
42. A. Thiruvengadam, S.L. Ruby, and M. Gunasekaran, Experimental and Analytical Investigations on Liquid Impact Erosion, STP 474, ASTM, 1970, p 249-287
43. F.J. Heymann, A Survey of Clues to the Relationship between Erosion Rate and Impact Parameters, *Proceedings of the Second Meersburg Conference on Rain Erosion and Allied Phenomena,* Royal Aircraft Establishment, England, 1974, p 683-760
44. F.J. Heymann, Toward Quantitative Prediction of Liquid Impact Erosion, STP 474, ASTM, 1970, p 212-248
45. G.F. Schmitt, Jr., Erosion Rate-Velocity Dependence for Materials at Supersonic Speeds, STP 474, ASTM, 1970, p 323-352
46. M.C. Rochester and J.H. Brunton, Influence of Physical Properties of the Liquid on the Erosion of Solids, STP 567, ASTM, 1974, p 128-151
47. R.H. Richman and W.P. McNaughton, Correlation of Cavitation Erosion Behavior with Mechanical Properties of Metals, *Wear,* Vol 140 (No. 1), 1990, p 63-82
48. G.S. Springer, *Erosion by Liquid Impact,* John Wiley & Sons, 1976
49. A. Karimi and W.R. Leo, Phenomenological Model for Cavitation Erosion Computation, *Mater. Sci. Eng.,* Vol 95, 1987, p 1-14
50. G. Beckmann and J. Krzyżanowski, A New Model of Droplet Impact Wear, *Proceedings of the Seventh International Conference on Erosion by Liquid and Solid Impact* (ELSI-VII), Cavendish Laboratory, University of Cambridge, England, 1987, p 12-1 to 12-7 [also, G. Beckmann, A Model of Erosion of Metals by Liquid Impact--I, *Kernenergie,* Vol 32 (No. 9), 1989, p 375-381; and A Model of Erosion of Metals by Liquid Impact--II, *Kernenergie,* Vol 32 (No. 12), 1989, p 471-476]
51. F.N. Mazandarany and F.G. Hammitt, "Russian and Eastern European Literature on Cavitation Resistant Steels," Report MMPP-344-4-T, University of Michigan, Aug 1969
52. A. Akhtar, A.S. Rao, and D. Kung, Cavitation Erosion of Stainless Steel, Nickel and Cobalt Alloy Weld Overlay Materials, *Proceedings of the Conference on Coatings and Bimetallics for Aggressive Environments,* American Society for Metals, 1985, p 125-142

53. R. Simoneau *et al.,* Cavitation Erosion and Deformation Mechanisms of Ni and Co Austenitic Stainless Steels, *Proceedings of the Seventh International Conference on Erosion by Liquid and Solid Impact* (ELSI-VII), Cavendish Laboratory, University of Cambridge, England, 1987, p 32-1 to 32-8 [also, R. Simoneau, A New Class of High Strain-Hardening Austenitic Stainless Steels to Fight Cavitation Erosion, *Proceedings of the 13th IAHR Symposium,* Montreal, 1986, p 83.1 to 83.20]

54. Letters in *ASTM Standardization News,* by F.J. Heymann in Vol 11 (No. 10), Oct 1983, p 7; by G.S. Springer in Vol 12 (No. 6), June 1984, p 10

55. P. Veerabhadra Rao and D.H. Buckley, "Empirical Relations for Cavitation and Liquid Impingement Erosion Processes," NASA Technical Paper 2339, Aug 1984 [also, P. Veerabhadra Rao, D.H. Buckley, and M. Matsumura, A Unified Relation for Cavitation Erosion, *Int. J. Mech. Sci.,* Vol 26 (No. 5), 1984, p 325-335

56. Z. Ruml and A. Liška, A Model for the Erosion of Steam Turbine Blade Materials, *Proceedings of the Seventh International Conference on Erosion by Liquid and Solid Impact* (ELSI-VII), Cavendish Laboratory, University of Cambridge, England, 1987, p 15-1 to 15-8

57. A.L. Shubenko and A.E. Kovalsky, On Prediction of Erosion Wear of Details (*sic*) on the Basis of Its Kinetic Model by Impact of Liquid Drop on Polydisperse Flows of Moisture, *Proceedings of the Seventh International Conference on Erosion by Liquid and Solid Impact* (ELSI-VII), Cavendish Laboratory, University of Cambridge, England, 1987, p 14-1 to 14-6

58. B. Staniša, M. Dičko, and K. Puklavec, Erosion Process on the Last Stage Rotor Blades of Turbines in Service, *Proceedings of the Seventh International Conference on Erosion by Liquid and Solid Impact* (ELSI-VII), Cavendish Laboratory, University of Cambridge, England, 1987, p 16-1 to 16-7

59. P.A. March, Evaluating the Relative Resistance of Materials to Cavitation Erosion: A Comparison of Cavitating Jet Results and Vibratory Results, *Proceedings of the ASME Cavitation and Multiphase Flow Forum,* FED-Vol 50, American Society of Mechanical Engineers, 1987, p 79-83

60. M. Johnson *et al.,* The Resistance of Nickel and Iron Aluminides to Cavitation Erosion and Abrasive Wear, *Wear,* Vol 140, 1990, p 279-289

Slurry Erosion

John E. Miller, White Rock Engineering, Inc.

Introduction

A SLURRY can be described as a mixture of solid particles in a liquid (usually water) of such a consistency that it can be readily pumped. The term "slurry erosion" is strictly defined as that type of wear, or loss of mass, that is experienced by a material exposed to a high-velocity stream of slurry. This erosion occurs either when the material moves at a certain velocity through the slurry or when the slurry moves past the material at a certain velocity.

Typical pumpable slurries possess inherent "apparent abrasivity," which must be determined by testing to enable cost predictions for pump replacement parts or other equipment used for slurries. Apparent abrasivity, without inhibition, is the complex synergistic reaction of many factors (Fig. 1). This reaction, known as the Morrison-Miller effect (Ref 1), is such that the wear response of a given material in a certain slurry does not indicate how that material would respond to another slurry. Similarly, the effect of a certain slurry on one material does not indicate how it would affect another material.

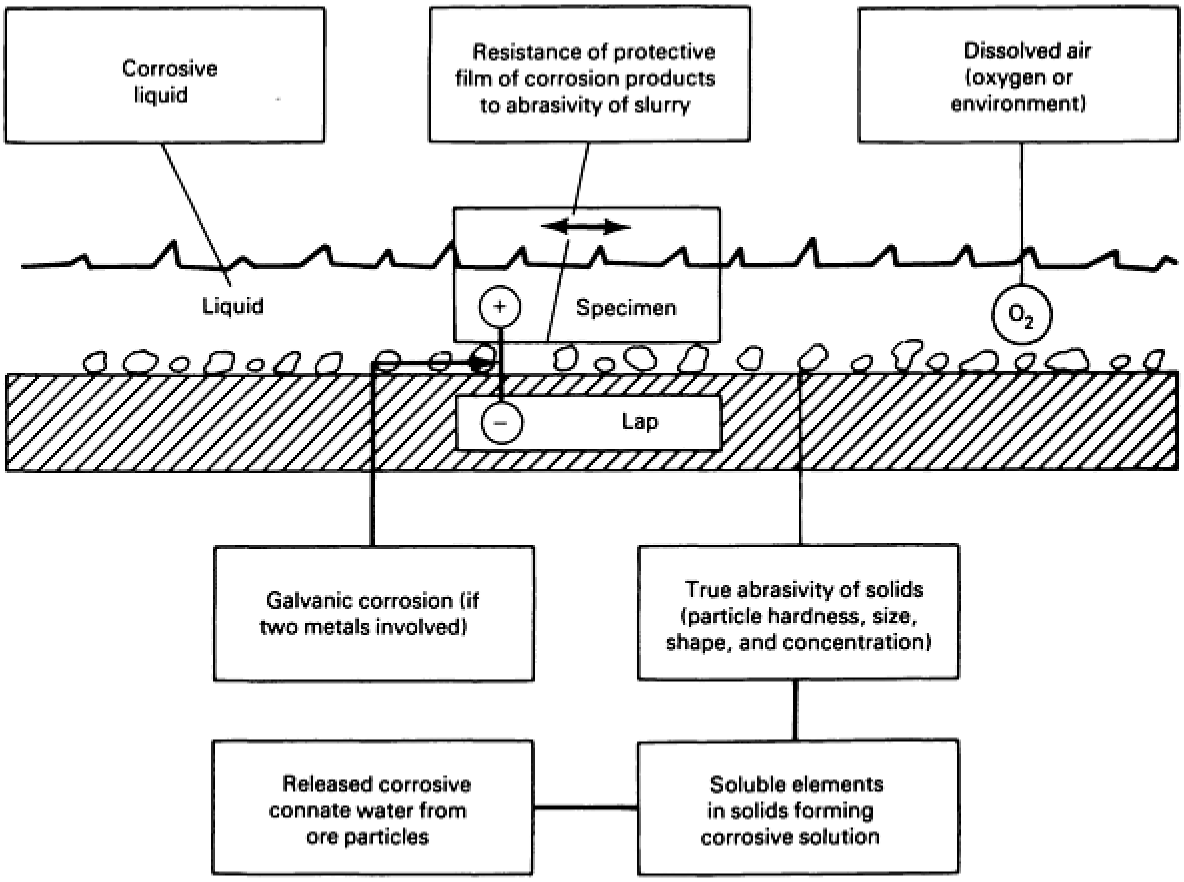


Fig. 1 Synergistic effects of seven factors in slurry abrasivity

Other modes of wear are also encountered when handling slurries, the most severe of which is the combination of abrasion and corrosion. The elusive erosion-corrosion combination was recognized as early as 1967. As shown in Fig. 2, it was found that the mass-loss rate was the reverse of what had been expected.

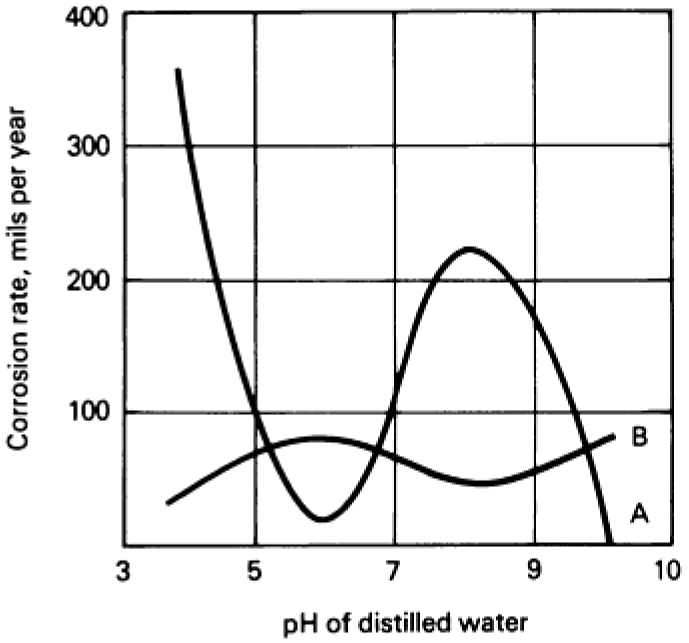


Fig. 2 Effect of pH of distilled water on erosion-corrosion of carbon steel at 50 °C (120 °F) and velocity of 12 m/s (39 ft/s). Curve A represents erosion-corrosion disk, and curve B represents specimens immersed in tank. Source: Ref 2

The slurry-handling process has been of increasing interest to industry, particularly because of the relatively new method of transporting minerals (coal) and other solids over long distances through pipelines at pressures in the order of 14 MPa (2 ksi). Therefore, this article identifies a variety of slurry wear modes, their effects, and the tests associated with them. Dry abrasivity is another wear mode that is only briefly discussed in this article because it is seldom, if ever, encountered in slurry handling.

Slurry Wear Modes

The most common wear modes are described below and illustrated in Fig. 3. Several of these modes will be characterized by:

- The types of motion to which they respond, that is, reciprocating, oscillating, circular, or continuous
- The slurry media, that is, oil or liquids other than water
- The ratio of areas of mating parts, as shown in Fig. 3

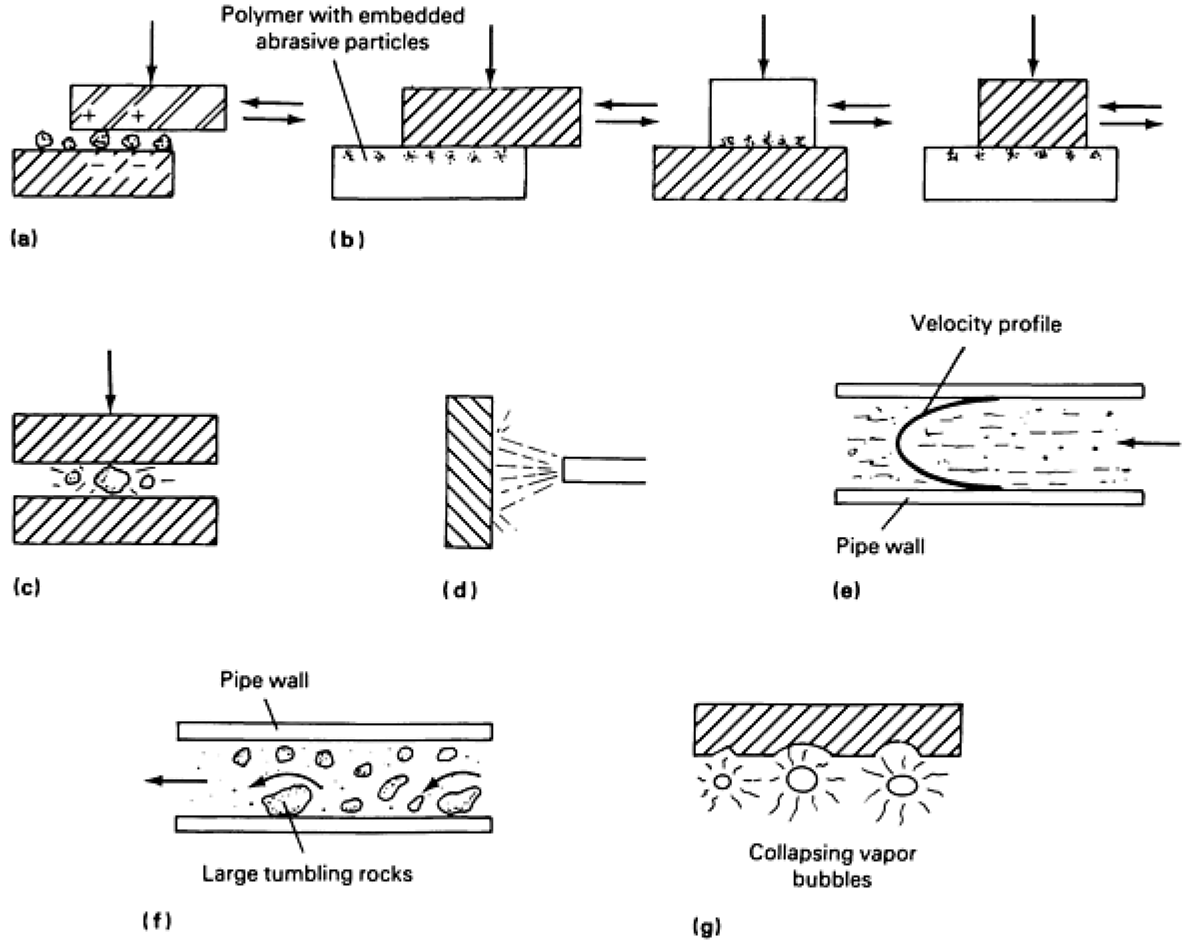


Fig. 3 Slurry erosion wear modes. (a) Abrasion-corrosion. (b) Scouring wear, with wear areas equal (left) and unequal (center and right). (c) Crushing and grinding. (d) High-velocity erosion. (e) Low-velocity erosion. (f) Saltation erosion. (g) Cavitation

Abrasion-corrosion wear is the result of any metal-to-metal rubbing in the presence of abrasive solids in a corrosive liquid. This is the most destructive and misunderstood mode encountered when handling slurry, aside from high-velocity erosion. Typical parts of reciprocating pumps that are involved are: metal-to-metal valves and seats (upon each closure), metal piston parts that rub on metal liners and plungers, or piston rods that rub against metal stuffing box parts or trim. In the case of parts of different metals exposed to slurry, electrolytic effects are included in this mode.

Scouring wear is encountered with elastomer-to-metal rubbing. Abrasive solid particles become embedded in the softer elastomer, or rubber, as in the case of pistons, packing, and valve inserts.

Crushing and grinding occur in abrasive metal-to-metal contact. An example is a valve that repeatedly closes with great force against solid particles that are trapped between the valve and seat at closure.

High-velocity erosion can become very destructive, although it is not the mode of wear that is usually found in reciprocating pumps handling slurry. High velocity is usually considered to be greater than 6 to 9 m/s (20 to 30 ft/s). When a valve seat or piston wears to the extent that a slight leak develops, the extremely high velocity of slurry leakage through that small gap can result in catastrophic and rapid failure of the parts, as well as the costly liquid end of the pump (Fig. 4 and 5). Slurry-throttling valves and parts that are downstream also experience this type of rapid wear, as do the impellers and cut-water mechanisms of centrifugal pumps.

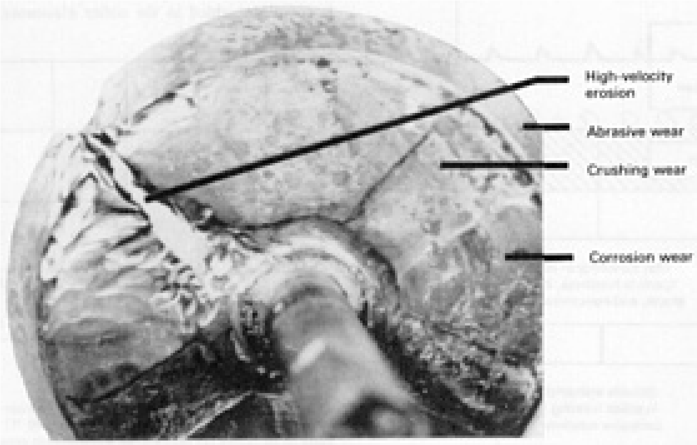


Fig. 4 Typical worn and washed-out valve

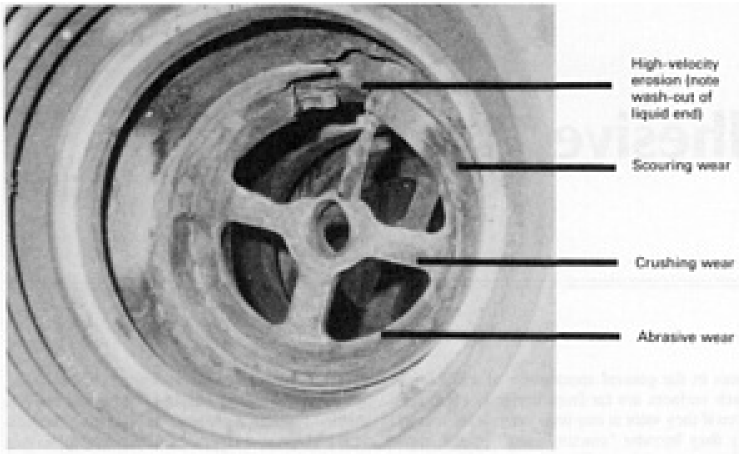


Fig. 5 Typical worn and washed-out slurry valve seat

Low-velocity erosion is usually a low-rate wear mode that occurs when there is a flow of slurry at regular low velocities. In a pipeline with laminar flow, the velocity profile (in the shape of a parabola) is such that the velocity near the wall of the pipe is nearly zero, and minimum wear takes place. The impellers and cut-water mechanism of centrifugal pumps are sometimes subjected to this mode.

Saltation wear results from the transport of a sediment, where the particles are moved forward in a series of short intermittent bounces from a bottom surface. Rapid wear of this type occurs, for example, in pipelines that transport phosphate rock from slurry pit to the processing plant. The pipe requires frequent turning to distribute the wear.

Cavitation can result in damage to the metal in the liquid end of the pump. The parts of a reciprocating pump can also sustain damage from the microscopic, but intense, liquid pressure blasts against the metal near the cavitation area, following the repeated collapse of the vapor bubbles.

Effects of Wear

The effect of solids concentration on abrasivity should be recognized. Figure 5 shows how abrasivity increases very rapidly from a concentration of 0 to about 10%, and then begins to flatten. It is interesting to note that the slurry that is well-known in the oil well drilling industry as "drilling mud" in itself is not abrasive. Rather, its abrasivity comes from the solid particles of crushed and broken rock, or cuttings, that are generated in the drilling process. Contaminated mud is usually "cleaned" by settling, screening, or being subjected to centrifuge. The curve in Fig. 6 is supported by the fact that

the sand content of drilling mud must be reduced to less than 2% before effective improvement in the life of pump parts can be realized.

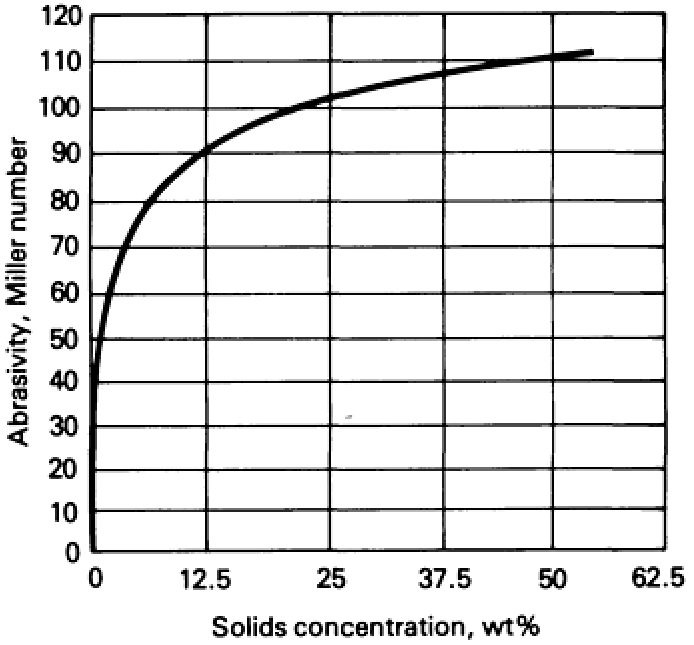


Fig. 6 Solids concentration versus abrasivity for 70 mesh urn sand, showing abrupt change in relationship of solids concentration to abrasivity in region below about 10 to 12% solids

ASTM G 75 Tests

Miller numbers are used to determine the abrasivity of slurries, based on the rate of metal loss from a standard 27% chrome-iron wear block that reciprocates through any slurry, on a rubber lap, with an imposed load of 22.2 N (5 lbf) placed on the wearing block. The higher the number, the greater the aggressive effect of the slurry on part life. The additional effect of corrosion (usually present in liquid slurries, even those mixed with distilled water) was slow to be recognized. This was because such chrome-iron is rather corrosion resistant and the original test actually fulfilled its objective to reveal the "true abrasivity" of the dry particles. The effects of both abrasion and corrosion must be recognized in the operation of any slurry-handling system.

Table 1 lists typical Miller numbers for selected slurry materials. The wide variation in Miller numbers for some materials is due to the inclusion of varying amounts of "tramp" materials that usually occur with the basic mineral. Information about the factors that contribute to Miller number abrasivity can be found in the ASTM G 75 standard.

Table 1 Typical Miller numbers for selected slurry materials

Material	Miller number(s)
Alundum (400 mesh)	241
Alundum (200 mesh)	1058
Aragonite	7
Ash	127
Ash, fly	83, 14
Bauxite	9, 33, 50, 76, 134
Calcium carbonate	14
Carbon	14, 16
Carborundum (220 mesh)	1284
Clay	34, 36
Coal	6, 10, 21, 28, 47, 57
Copper concentrate	19, 37, 58, 68, 111, 128

Detergent	6, 8
Dust, blast furnace	57
Gilsonite	10
Gypsum	41
Iron ore (or concentrate)	28, 37, 64, 79, 122, 157, 234
Kaolin	7, 30
Lignite	14
Limestone	22, 30, 39, 43, 46
Limonite	113
Magnesium hydrate	4
Magnetite	64, 71, 134
Microsphorite	76
Mud, drilling	10
Nickel	31
Phosphate	68, 74, 84, 134
Potash	1, 2
Pyrite	194
Quartzite	99
Rutile	10
Salt brine	11
Sand and sand fill	51, 68, 85, 116, 138, 149, 246
Sea bottom	11
Shale	53, 59
Serpentine	134
Sewage, digested	15
Sewage, raw	25
Sodium sulfate	4
Soda ash tailings	27
Sulfur	1
Tailings (all types)	24, 61, 91, 159, 217, 480, 644
Tar sand	70
Waste, nickel	53
Waste, coal	22, 28

Slurry Abrasion Response (SAR). It soon became evident that the response of various metals differed widely in different slurries. Therefore, a new test, the Miller number "in reverse," was developed. The slurry abrasion response (designated by an SAR number) of any selected material in any given slurry is based on the same Miller number procedure. It was demonstrated that the aspects of corrosion varied widely with different materials and interfered or sometimes obscured the effects of true abrasive erosion, or wear.

Dry Abrasivity

As a result of further work, it was decided that all Miller or SAR number tests would be run in duplication. One would have the "as-received" material, and the other would have an inhibited slurry obtained by adding a dose of strong NaOH solution to raise the pH to the order of about 13. Thus, a "number" approaching the dry abrasivity could be obtained by the reduction of corrosion effects. An increase in the pH of the slurry by the addition of NaOH (as is sometimes done in actual systems) reduces the Miller or SAR number, in most cases.

References

1. Morrison and Corbett, "A Modified G-75 Abrasion Test for Corrosive Environments," Corrosion Testing Laboratories, Inc.
2. M.G. Fontana and N.D. Greene, *Corrosion Engineering,* McGraw-Hill, 1967

Sliding and Adhesive Wear

K.C. Ludema, University of Michigan

Introduction

SLIDING AND ADHESIVE WEAR refer to a type of wear generated by the sliding of one solid surface along another surface. Erosion, cavitation, rolling contact, abrasion, oxidative wear, fretting, and corrosion are traditionally excluded from the class of "sliding" wear problems even though some sliding (slip) may occur in some of these types of wear. Apparently, sliding wear is a category of wear that is "left over" when all other types of wear are identified under separate headings.

Adhesive wear is as ambiguously defined as is sliding wear, though the two are not strictly synonymous. Adhesive wear denotes a wearing action in which no specific agency can be identified as the cause of the wear. Adhesion has been confirmed in the 1930s to be a major contributor to sliding resistance (friction) and was inferred in mechanics at least to be operative in wear as well. Thus, if no abrasive substances are found, if the amplitude of sliding is greater than that in fretting, and if the rate of material loss is not governed by the principles of oxidation, and so on, adhesive wear is said to occur.

Adhesion is most clearly demonstrated in sliding systems when, for example, a shaft seizes in a bearing. Again, adhesion is implicated in galling and scuffing (which some researchers classify as surface damage rather than wear) even when no material is lost from the system. From these examples, it can be seen that adhesion is not the cause of wear, but only the consequence of contact. Wearing occurs when interfaces in contact are made to slide and the locally adhered regions must separate. This separation may occur by one or two of the failure modes of solids, resulting in a very wide range of wear rate.

The purpose in defining and studying specific wear modes is to find ways to prolong the wear life of sliding systems. Each mode of wear can best be resisted by the proper choice of material, lubricant, and method of operation. Unfortunately, the making of slow-wearing products requires considerable experience, because there are very few fundamental principles in friction and wear to use in the design process.

Sliding Surface

The focus of this section is the case of long-distance sliding of nominally clean and dry surfaces in the general atmosphere of technology. Such surfaces are far from perfectly clean, and even if they were at one time very clean, inevitably they become "contaminated" by the atmosphere around them. We may thus define "standard conditions" for sliding, for the three classes of materials (metals, polymers, and ceramics) as follows:

- Metals will usually be covered with a film of oxide, which is covered by a second film of adsorbed gases and hydrocarbons (oils)
- Polymer surfaces may contain absorbed water and "bloomed" plasticizers, and will likely be covered with mold release agents, adsorbed gases, and oils
- Ceramic materials will also be covered with adsorbed gases and oils; in addition, the nonoxide ceramics may oxidize

The thickness of the adsorbed layers and coatings is of the order of 10 nm (100 Å). Such thin films are not discernible by optical means or by scanning electron microscopy (SEM) techniques. The presence of thin films must therefore be verified by sophisticated chemical analytical instruments, or taken on faith. These films are very important because they intervene in the bonding or adhesion between the substrate materials. Variations in the thickness and properties of these films is one major reason for the very wide range of friction data reported for all materials.

These films control the variations in wear as well. The reason that these films control wear becomes evident in the example that follows. When a piece of brass "contacts" a piece of silicon carbide, the weight of the upper body is carried through a chain of bonds, from the silicon carbide to a layer of silicon dioxide, to two layers of adsorbed gas, to a layer of ($CuO + ZnO_2$), and finally to the brass. The contact bridge between the brass and silicon carbide is shown in Fig. 1.

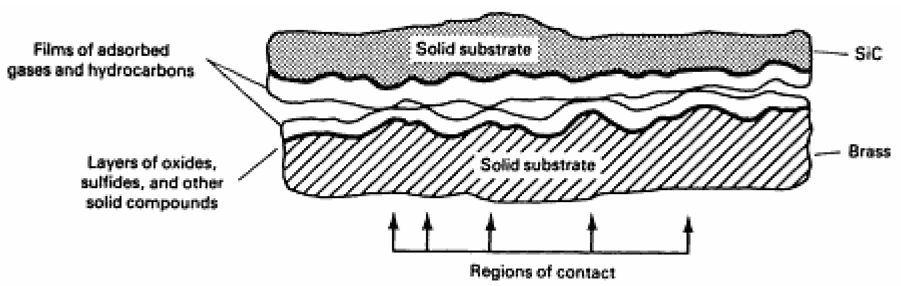


Fig. 1 Schematic of a bond bridge produced when two solid surfaces are in contact with each other. It should be noted that when two rough surfaces are brought together, actual contact occurs only in a few isolated regions.

When two such surfaces slide against each other, there will be some squeezing out of the adsorbed films, producing some contact between solid (oxide, sulfide, and so on) layers. If the contact pressure is high, a solid layer may fracture and expose substrate material, particularly if the substrate beneath the solid layer deforms plastically. Small regions of both substrate materials may finally come into contact and bond together.

Material-Dependent Bond Strength. At the small points of substrate-to-substrate contact, the bond strength will depend on the materials. Identical metals (for example, iron against iron) bond together most completely when the bond strength is very near the strength of the metals themselves. Dissimilar metals will usually bond with less strength in the first instant, but these "bonds" increase in strength with time, particularly between metal pairs that will mix or alloy via diffusion. Two pieces of identical ionic or covalent atomic structures, such as ceramic materials, will bond very well if the atomic lattices are well aligned. Bond strength will usually be less where the lattice dimension of one is different from that of the other. Recall that oxides, sulfides, and other compounds are also ionic and covalent structures. Finally, the atoms in polymer chains are usually bonded covalently. The polymer chains bond to each other and to other material by the weaker van der Waals bonds.

Conformity of Contacting Surfaces. A second consideration in bonding is the conformity of the contacting surfaces. Ductile materials will conform somewhat via plastic flow, but the brittle materials will not conform.

Sliding adds shear stress to any normal stresses already imposed, increasing the probability of plastic flow and fracture of materials in the contact region. After some sliding, there will be a layer of turbulently mixed substances, some of which will fall out as wear debris, but most of which will remain as a "transfer film." The wear rate will depend on the properties of all of the substances, the properties of the mixed composite material, and the tendency for bits of that turbulently mixed material to leave the system. The latter tendency will depend on the shape and vibration modes of the contacting bodies, the temperature, possible chemical reactivity with the surrounding environment, and many other parameters. It is clear that adhesion is prominent, although not well-defined in these events.

Wear Equations, Design Criteria, and Material Selection

The above brief description of the wearing process showed it involves very many variables, few of which have been researched sufficiently to formulate wear equations that can be used by designers. Some wear equations can be found in the literature, but none can be generally applied. The most widely publicized equation is due to Archard, in which the rate of wear, Υ (volume per unit of time), is:

$$\Upsilon = k \cdot \frac{WV}{H} \quad \text{(Eq 1)}$$

where W is the applied load, V is the sliding speed, H is the hardness of the softest of a pair of the materials, and k is a constant sometimes referred to as a wear coefficient.

Clearly, neither this equation nor any other published equation can adequately describe what occurs in Fig. 1. However, Eq 1 is widely known, and values of k for various materials are increasingly found in the literature. These numbers range from 10^{-3} for unacceptable wear rates to $\leq 10^{-7}$ for practical systems. These numbers serve as very broad guidelines for anticipated wear rates of materials, just as do published values of coefficients of friction. However, designers need coefficient values of high reliability within a range of ±20% or better in order to design acceptable systems.

Most engineers are certain that one or more wear equations exist that are more applicable to their problem than is suggested above. This belief likely arises from the existence of good equations in most other topics in engineering; there are equations for the bending of beams, the electrical resistance of materials, the rate of heat transfer, and the natural frequency of vibration of machinery. These are relatively simple phenomena, however, governed by very few variables. Adhesive wear, on the other hand, appears to involve over 25 variables, some of which can be measured easily and others of which are difficult or impossible to measure.

The point should be emphasized that there is no equation which can be generally applied. In fact, there are many practical situations in which the published equations lead to erroneous conclusions. For example, cam shafts in auto engines are not fully hardened because in that state they wear far too fast; some brake materials wear faster at low speeds than at high speeds; and some hardenable steels wear faster at low sliding speed than at high sliding speed. It is therefore important not to rely on any specific equation without considerable simulative testing. In testing, it is very important to characterize the sliding system thoroughly, including factors such as stiffness of the system, vibration modes, and atmospheres surrounding the test. Finally, even the characterization of the system as done by humans is inadequate. The most accurate method of determining whether a laboratory test truly simulates a mechanical component under study is to verify that the progression of surface change (wear rate, friction, appearance, and so on) and the wear debris in both are identical or nearly identical.

The sections that follow in this article are descriptions of the wearing of metals, polymers, and ceramic materials. These are presented to provide a perspective on the complexity of dry sliding wear, also frequently referred to as adhesive wear, conditions.

Primary Parameters in the Wear of Metals, Polymers, and Ceramics

Material selection for wear resistance requires careful consideration of the operating environment of the workpiece in addition to the total functional performance required of the workpiece itself. Wear properties of the steels vary widely with processing and heat treatment. Polymers are selected for sliding contact applications because of inherent properties such as inertness to many chemicals, relatively low galling tendency, and self-lubricating properties. Ceramics are used where extreme resistance to high-temperature oxidation or resistance to highly corrosive materials or gases is required.

Dry Sliding Wear of Clean Metals

The term dry means no deliberate lubrication, and the term clean indicates no obvious oxide scale or greasy residue.

60-40 Brass versus Tool Steel. A.W.J. DeGee and J.H. Zaat (Ref 1) found that adhesive wear occurs via two different mechanisms. These two mechanisms produce two effects, which are shown in Fig. 2 for 60-40 brass rubbing against tool steel. Brass was found to have transferred to steel where it remains attached, but some brass is also removed (worn) from the system. The extent of each event depends on the zinc content in the brass:

- *Local adherence of brass to steel:* No iron is observed to be present in the wear fragments. Brass particles come loose from the steel, but new material fills the impression again. Most of the steel surface remains undisturbed as seen by the unaltered surface features. The oxide film present on the brass is composed of CuO. The combination of CuO plus iron oxide apparently lubricates well except at some isolated points where brass transfers to steel
- *Continuous film:* The oxide present on the brass surface is zinc oxide. This oxide possibly does not lubricate and there is a considerable adhesion area. A thin film of brass is found on the steel in which the wear particles are large but few. This film covers (that is, fills) the surface roughness so that the

mechanism is independent of surface finish

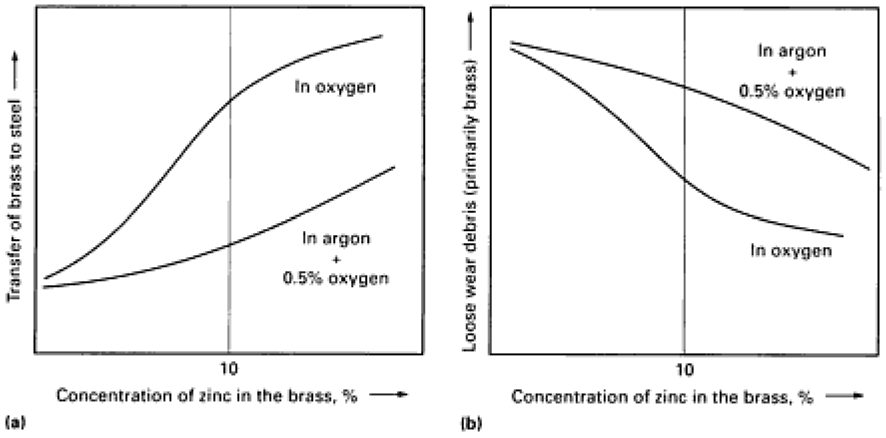


Fig. 2 Two different mechanisms that generate adhesive wear when 60-40 brass is rubbed against tool steel. (a) Transfer of brass to steel. In an oxygen atmosphere, the transfer film is mostly CuO, locally adhering, and dependent on surface roughness. In an argon/0.5% oxygen atmosphere, the continuous transfer film consists primarily of ZnO. (b) Removal of brass from system due to loose wear debris. Source: Ref 1

Lancaster (Ref 2) also studied the effect of 60-40 brass on high-speed steel and concluded:

- Transition between severe wear and mild wear is affected by sliding speed, applied load, and ambient temperature (Fig. 3 and 4; note the effect of atmosphere)
- Thickness of the oxide is the critical factor. The actual thickness is determined by the time available to reoxidize a denuded region and by the rate of formation of the oxide (Fig. 5). Reoxidation time is determined by sliding speed only. Oxide formation rate is influenced by temperature
- Hard metals do not necessarily wear soft metals away (Fig. 6)
- The ratio of various elements in the wear products is not predictable (Fig. 7)
- Wear rate, Υ, coefficient of friction, μ; and electrical contact resistance can vary independently (Fig. 8)

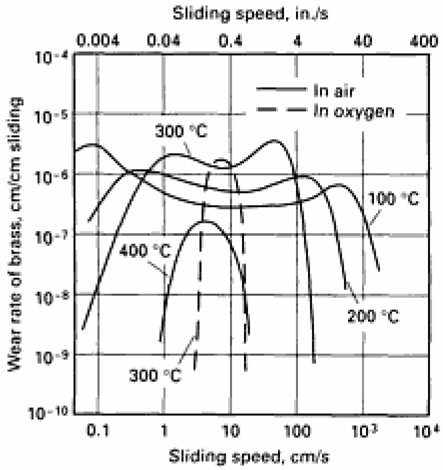


Fig. 3 Plot of wear rate versus sliding speed for a 60-40 brass on high-speed tool steel system using a 3 kgf (7 lbf) load both in air and in an oxygen atmosphere

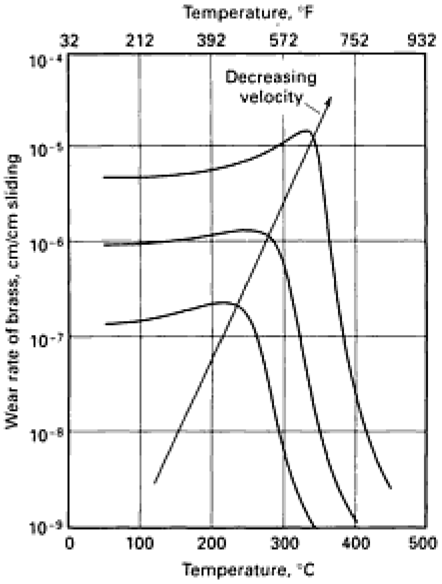


Fig. 4 Plot of wear rate versus ambient temperature as a function of sliding speed for 60-40 brass being rubbed on high-speed tool steel

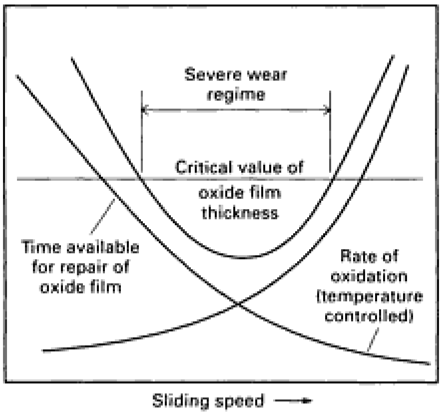


Fig. 5 Effect of competing factors on oxide film thickness when 60-40 brass is rubbed by high-speed tool steel

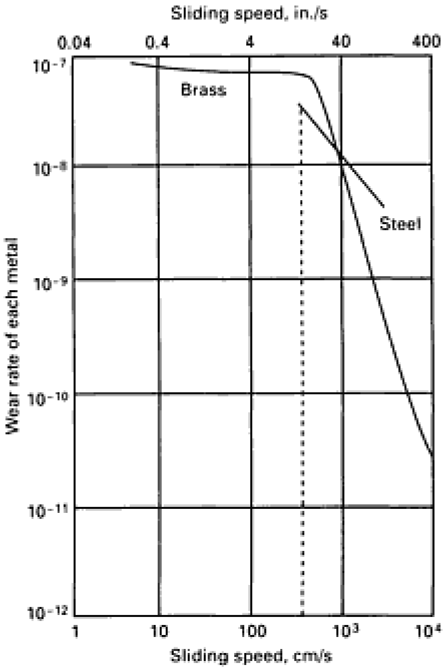


Fig. 6 Plot of wear rate versus sliding speed for 60-40 brass rubbed with high-speed tool steel in which brass hardness differs from steel hardness

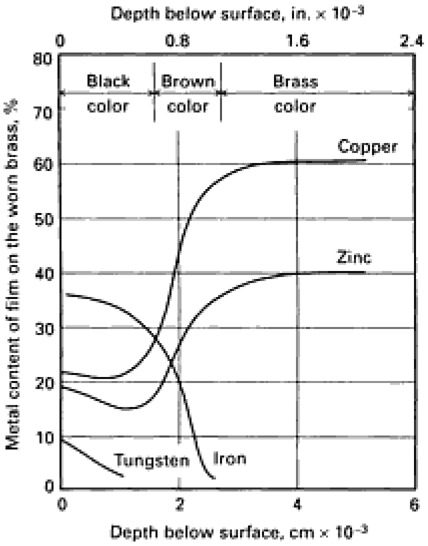


Fig. 7 Surface composition of a worn 60-40 brass pin rubbed by an 18W-4Cr-1V-1C composition high-speed tool steel wheel

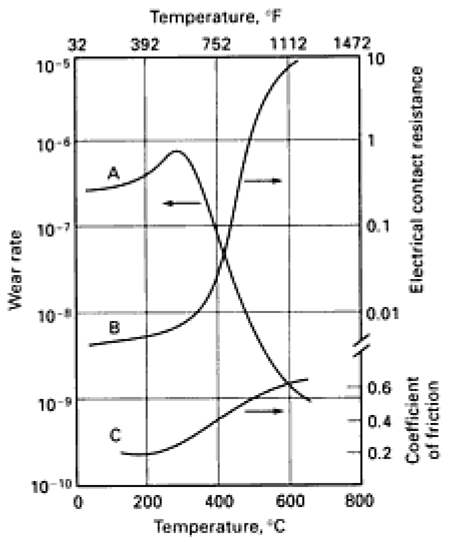


Fig. 8 Friction, wear, and contact resistance as a function of temperature for a 60-40 brass/high-speed tool steel system. A, wear rate, Υ; B, electrical contact resistance; C, coefficient of friction, μ

0.12% Carbon Steels. Low-carbon steel is a readily available low-cost material capable of being processed in numerous ways to resist wear. N.C. Welsh (Ref 3) tested 0.12% C steels and came to the following conclusions:

- Increasing applied load may decrease wear rate (Fig. 9), presumably by increasing the temperature at the points of contact. The rubbing temperature becomes high enough to form austenite. Partial proof of this transformation can be seen in Fig. 10
- Wear debris from rubbing of steel parts is a mixture of oxides (Fig. 11)

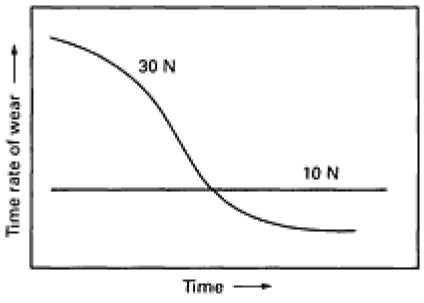


Fig. 9 Plot of wear rate versus time for a low-carbon (0.12% C) steel tested with a 10 and 30 N (1 and 3 kgf) load

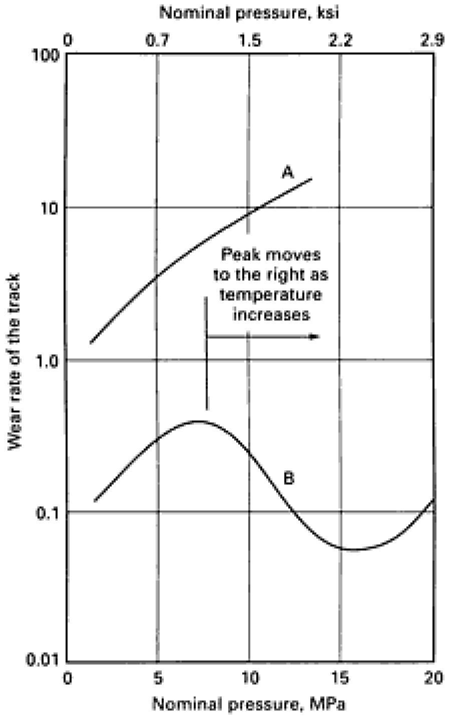


Fig. 10 Plot of wear rate versus nominal pressure as a function of unhardenable and hardenable 0.12% C steels. A, wear with galling (soft steel rubbing on soft steel); B, soft steel rubbing on hardenable steel

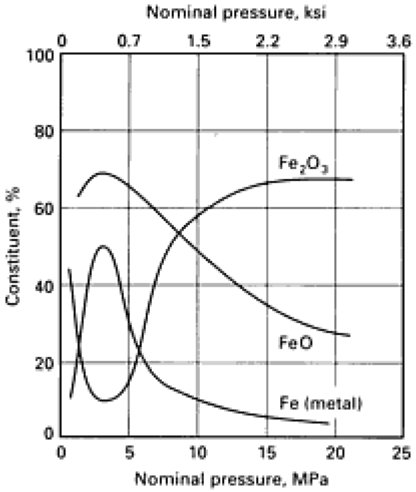


Fig. 11 Plot of wear debris constituent concentration versus nominal pressure for the conditions (debris from soft 0.12% C steel sliding on hardenable 0.12% C steel) in Fig. 9

N.C. Welsh (Ref 4) later measured wear rate versus load for steel on steel and again found transitions between severe wear and mild wear. His data were published in the form shown in Fig. 12. The data indicated that:

- Effect of hardness is to diminish the likelihood of transition to severe wear. It can be speculated that the critical oxide thickness is less for hard substrates than for the soft substrate
- Transition does not occur immediately with a change in sliding conditions
- Low sliding speed shifts the transition to the right and high speed shifts the transition to the left
- Inert gas broadens the transition region, whereas oxygen narrows the transition region
- In the mild wear regime, initial wear rate is high where the oxide is chemically removed and rubbing

resumes (Fig. 13)

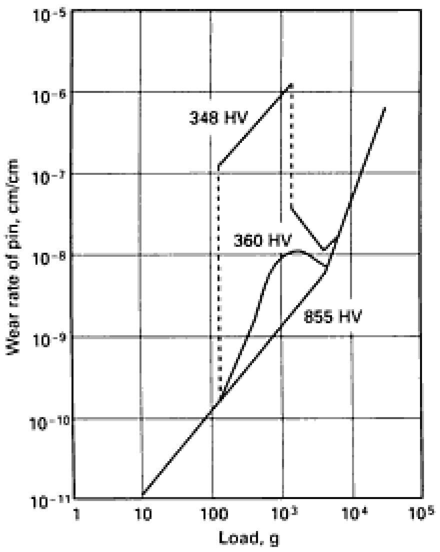


Fig. 12 Plot of wear rate versus load for three 1050 medium carbon steels of different hardness (348, 360 and 855 HV)

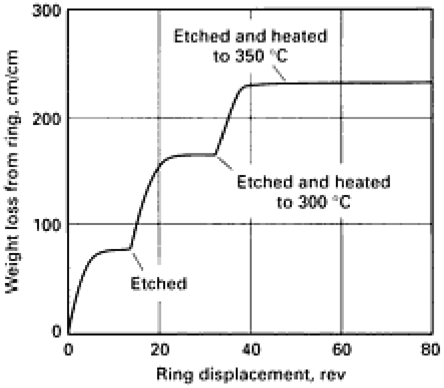


Fig. 13 Plot of weight loss from ring versus ring displacement to show effect of heating steel on wear rate. Velocity, 1 m/s (3.3 ft/s)

Dry Sliding Wear of Polymers

In the rubbing of polymers (and also of dry bearing materials that include polymers) on metals, a very thin film of polymer, invisible to the naked eye, is first established upon the metal surface; the presence, and indeed the nature, of this thin transferred film can play a crucial role in the subsequent wear behavior.

Wear of Metals Generated by Plastic. The friction of polymers is about the same as that of metals (with the exception of PTFE), but the seizure resistance of plastics is superior to that of soft metals. General uncertainty exists about the effect of surface roughness on wear rate, and the astounding finding has been made that some polymers wear metals away without the presence of abrasives.

The general state of understanding of polymer wear is that rubbing surfaces experience a break-in period, followed by a steady-state behavior, which is often referred to as linear wear (Fig. 14).

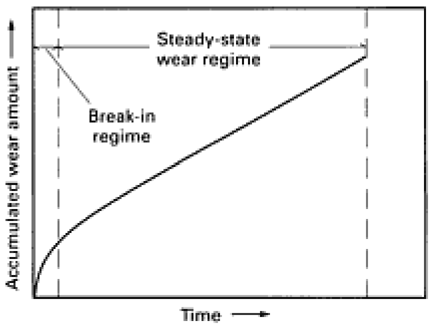


Fig. 14 Typical wear behavior of a polymer, showing the break-in period and the steady-state period of the curve generated when wear is plotted versus time

It is in the linear region that people have been searching for a useful wear coefficient. A second number often used in connection with polymers (and bearing bushings) is a descriptor of the rubbing severity, above which severe or catastrophic wear may occur. The most widely known descriptor is the *PV* limit, where *P* is the average contact pressure (in psi) and *V* is the sliding speed (in ft/min). Each polymer has a unique *PV* limit as measured by some test, most often a washer test. The *PV* limit for polymers is often thought to be based on thermal behavior of the materials, but this is not a valid assumption.

Transfer films of polymer control the wear process. If a film of polymer is transformed on the metal counterface and remains firmly attached, the loss of the polymer from the system is minimal. If, during sliding, a particle of polymer is removed from the polymer bulk but does not remain attached to the metal, it is lost from the system. An intermediate state of wear is the case where a transfer film is formed, but fragments of the film are later lost, probably due to fatigue or some other mechanisms. Because the behavior of the system is very different in each of the three regimes, the discussion will now be focused in turn upon the break-in period, the steady-state wear regime, and the severe wear regime.

Surface Finish Orientation Relative to Sliding Direction. In the early stages of sliding, the rate or buildup of the transfer film is dependent on the orientation of surface finish relative to the sliding direction, and the rate varies depending on the type of contaminant or dirt on the sliding surface. Surface finish appears to have almostno effect on the steady-state wear rate, probably because the products of wear fill the grooves in the surface.

To check the effect of surface roughness on wear in the early stages of sliding, several polymers were slid on carbon steel with surface roughnesses from 0.1 to 3 μm (40 to 125 μin.) root-mean-square roughness (R_q), some parallel with the sliding direction and some perpendicular. For nylon 6/6 pins of 12 mm ($\frac{15}{32}$ in.) diameter with 100 N (10 kgf) load on each pin and a speed of 0.4 m/s (1.3 ft/s), the data in Fig. 15 were obtained. It appears that the nonlinear period, or beak-in period, may persist twice as long with parallel sliding than with perpendicular sliding and that the weight loss at the end of break-in may vary by a factor of four or more.

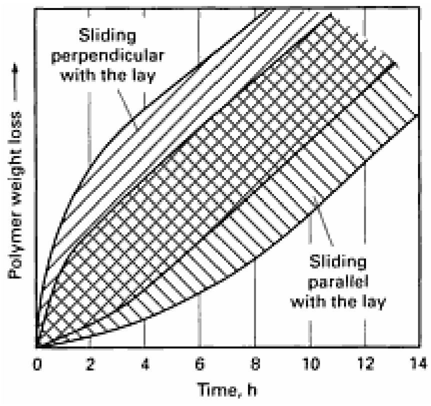


Fig. 15 Wear rate of nylon 6/6 polymer pin that was slid on rough (0.1 to 3 μm, or 40 to 125 μin., R_q roughness) carbon steel both parallel and perpendicular to the sliding direction. Test parameters: pin diameter, 12 mm ($\frac{15}{32}$ in.); load on pin, 100 N (10 kgf); velocity, 0.4 m/s (1.3 ft/s)

The break-in period designates the time when a film of polymer is transferred to the metal. The equilibrium film thickness for all tests run on various surfaces at one speed and one load were about the same.

Surface Cleanliness. The second factor in controlling the establishment of the transfer film is surface cleanliness. The tests were conducted with metal surfaces subjected to three conditions:

- Laboratory clean (covered with an adsorbed water film)
- Covered with a thin film of inert hydrocarbon (vacuum pump oil)
- Covered with a machine oil

In all cases, a transfer film begins to form and it may become continuous, each in a different time. The higher the temperature of the counter surface, the more quickly the transfer film formed.

As the transfer film forms, the loss rate from the polymer is high, but after the film forms, the wear rate is much lower (sometimes <1% of the initial rate). Thus, predicting the total wear rate of a bearing over some specified time should not be done on the basis of the steady-state wear rate. For short-time use of a polymer bearing, the break-in stage could produce much more wear than an amount based on predictions from data for steady-state wear. Furthermore, wear rate predictions are complicated by variations in temperature, by variations in the amount of contamination, by variations in speed, by start-stop or forward-reverse cycles, and by other factors.

The transition to severe wear occurs by the following sequence of events. During low wear rate sliding conditions, the transfer film remains as a transfer film behind the slider and provides (or becomes) a lubricant film upon which the slider rides in during later passes. If the temperature of the transfer film is high (such that the film of plastic has low viscosity) and if the low viscosity plastic does not wet the metal surface, the plastic agglomerates into spheres that are removed by the next slider that passes by (Fig. 16). Thus, the sliders are deprived of a lubricant film and instead give up some material to establish a new film that again is quickly detached. The difference between a tenacious and fleeting transfer film is a wear rate that varies by more than 2000 times. Relevant variables in this entire process must include the temperature and thermal properties (mass, diffusivity, and so on) of the counter surface and the contact area between the sliding parts.

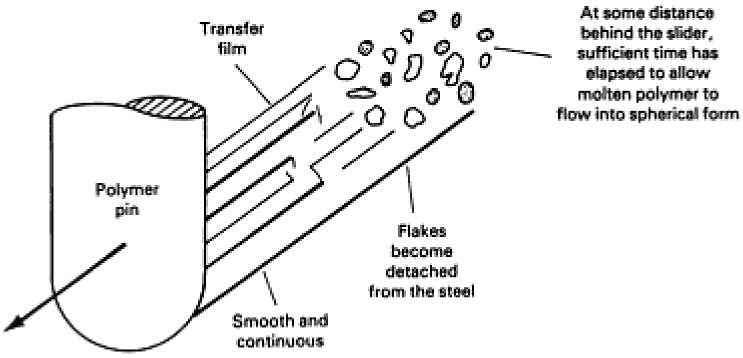


Fig. 16 Schematic showing mechanism that determines wear in a polymer moving at a high sliding speed

The overall effect of these mechanisms of material transfer and loss is the wear rate shown in Fig. 17. The wear rate increases with sliding severity, which is some combination of *P, V,* and other variables. The wear rate shown in Fig. 17 is most likely due to some combination of inadequate attachment of transfer film to the substrate and the removal of transfer film by fatiguing and other failure modes. If the surfaces are clean and attachment strength is high, at high severity of

sliding the transfer film does not readily fragment by fatigue ad wear is low. At high severity, the transfer film, however firmly or weakly attached, agglomerates and is lost as wear.

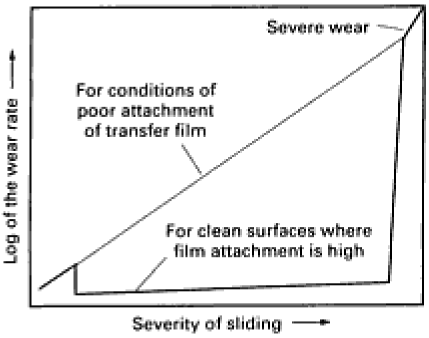


Fig. 17 Wear rate of polymers in both clean environments and contaminated environments

Metal Wear Generated by Plastics. When sliding some of the harder plastics on type 440C stainless steel that is hardened to 50 HRC, at all speeds and loads, iron and chromium are found attached to the polymer at the end of the test in the mild wear regime. These results indicate:

- A transfer film is not laid down by successive and simple shear from the polymeric sliding. Rather, there is a considerable turbulence or rolling of polymer within the transfer film, at least in the early life of the transfer film
- Transfer of iron and chromium to the polymer indicates that the metal is wearing away. Soft plastics did not wear away the type 440C stainless steel

Table 1 is a list of plastics used in one set of experiments with a sliding speed of 1.47 m/s (4.83 ft/s) and a load of 222 N (22.6 kgf) per pin of 13 mm ($\frac{1}{2}$ in.) diameter.

Table 1 Wear properties of selected polymers

Polymer type	Material	Shear strength		Coefficient of friction, μ	Effect of polymer on wear of type 440C stainless steel surface
		MPa	**ksi**		
Polyamide (PA)	Nylon 6/6	70.5	10.2	0.66	[(a)]
Polyoxymethylene (POM)	Delrin	65.5	9.5	0.65	[(a)]
High-density polyethylene (HDPE)		63.4	9.2	. . .	[(b)]
Polyoxymethylene + polytetrafluoroethylene (POM + PTFE) composite	Delrin AF	55.2	8.0	0.20	[(b)]
Polyamide (PA)	Nylon 11	41.4	6.0	0.50	[(b)]

(a) Hard plastic wears away steel surface.

(b) Soft plastic does not produce wear on steel.

Rubber is a polymer, but it differs from plastics in that its molecules are cross-linked. Thus, the migration or flow of the molecules is severely limited. Rubber wears by two mechanisms, tearing and fatigue. Ultimately, these are not very distinct mechanisms because tearing, or fracture, is failure in $\frac{1}{4}$ cycle of fatigue. However, the tearing mechanism is immediately visible. It occurs on rough, hard surfaces, particularly on a surface of sharp stones or abrasive paper. This

conclusion is supported by the observation that the ranking of several rubbers in a test on abrasive surfaces was the same as in tensile tests of the rubber. Even the temperature and rate dependency were the same.

The fatigue mechanism occurs when rubber slides on smoother surfaces. This mechanism was proposed on the strength of a correlation between the distance of sliding (number of deformation cycles due to passing asperities) until surface failure and the number of strain cycles to tensile failure; it is supported by a parabolic relationship between the applied stress and the fatigue life in both tests. The wear rates of various rubbers cover a very wide range. Generally, rubbers can be ranked within the range of wear rate by comparing the imposed stress state with the strengths (tearing and fatigue). The local coefficient of friction must be taken as one of the important causes of imposed stress.

Wear of Ceramic Materials

In general, ceramic materials are more brittle than metals and are often presumed to be incapable of plastic deformation. In addition, there are strong statements in the technical literature to the effect that ceramic materials are chemically inert and therefore a chemical component is not to be expected in the wearing of ceramics. The result has been that the technical community expects that the wear of ceramic materials will not be similar to the wear of metals. There are indeed differences, which are useful to list:

- Chemical reaction of a ceramic material with the environment does occur, often forming new chemical compounds, such as oxides (except on the ceramics that are oxides such as Al_2O_3, ZrO_2, and others), sulfides, phosphides, and others; these compounds influence the coefficient of friction just as they do in metals. These new surface compounds are often readily removed as wear products.
- Some ceramic materials (for example, Si_3N_4), are more ductile than some untempered martensite. In fact, white cast iron, although it is regarded as a metal, may contain well over 50% carbides of iron and chromium; these carbides are ceramic materials. However, such ceramic materials as silicon carbide and alumina are clearly more brittle than metals
- Brittle materials are much more prone to early failure due to cyclic stressing. In mechanical engineering, a response to cyclic stress is described as a fatigue property, but this term has a more technical meaning in material science. Clearly, though, the very brittle ceramics disintegrate quickly in cyclic stressing. One manifestation of this behavior may be seen in the pin-on-disk wear test. With pairs of ductile metals, the pin wears faster than does the disk; with pairs of brittle ceramic material, the disk wears faster than does the pin. The reason is that a position on the disk has the pin sliding over it repeatedly, imposing a cyclic stress on that position.
- The type and extent of failure of a surface (wear) depends strongly on the process used to form and prepare the surface. Severe processing induces fractures between grains and in the grains, so that when a surface is placed in sliding service, the fragmented surface wears away quickly. Wear particles of ceramics then carve out other grains and the wear is perpetuated.
- In repeat-pass sliding, some wear debris appears to be crushed and reattaches very firmly as a transfer film, which affects (often reduces) the wear rate by as much as a factor of 100. This reattachment appears to be interrupted by the presence of water and other fluids, which increases the wear rate by as much as tenfold

Prevention of Adhesive Wear

The following guidelines are recommendations to prevent adhesive wear in metals, polymers, and ceramics:

- Avoid sliding similar materials together, particularly metals
- If fatigue due to repeated high-contact pressure is not likely to be a problem, then high hardness is a desired property. However, avoid sliding hard metals against hard metals in lubricated systems to avoid scuffing and to accommodate debris
- Consider the effect of relative hardness of phases in materials. For example, a high-chromium cast iron may have a hardness of 400 HB, which is moderate. However, that cast iron may contain Cr_7C_3, which has a hardness of about four times that of 400 HB and will damage the counter surface considerably. The same applies to polymers, which seem rather soft relative to metals. However, wear-resisting

polymers often contain glass or some other hard filler that wears metal counterfaces rather severely. Hard phases in one body may fragment and become embedded in the counterface, which causes abrasion if the fragments extend above the surface.

- Even if done inadequately, lubrication will reduce wear. Some lubrication can be applied by providing an atmosphere that is corrosive in order to form surface films, many of which produce lower friction than if that film were not present

References

1. A.W.J. DeGee and J.H. Zaat, *Wear,* Vol 5, 1962, p 257
2. J.K. Lancaster, *Proc. R. Soc. (London) A,* Vol 273, 1963, p 466
3. N.C. Welsh, *J. Appl. Phys.,* Vol 28, 1957, p 960
4. N.C. Welsh, *Philos. Trans. R. Soc. (London) A,* Vol 257, 1958, p 1077

Fretting Wear

R.B. Waterhouse, University of Nottingham (United Kingdom)

Introduction

FRETTING is the small-amplitude oscillatory movement that may occur between contacting surfaces, which are usually nominally at rest. One of the immediate consequences of the process in normal atmospheric conditions is the production of oxide debris, hence the term "fretting wear" or "fretting corrosion" is applied to the phenomenon. The movement is usually the result of external vibration, but in many cases it is the consequence of one of the members of the contact being subjected to a cyclic stress (that is, fatigue), which gives rise to another and usually more damaging aspect of fretting, namely the early initiation of fatigue cracks. This is termed "fretting fatigue" or "contact fatigue." Fatigue cracks may also be initiated where the contacting surfaces are under a very heavy normal load or where there is a static tensile stress in one of the surfaces. There are cases where the movement is not simply tangential, but is complicated by the normal force also oscillating to the extent that the surfaces lose contact in each cycle. This leads to a hammering effect, which is termed "impact fatigue." In this case, the phase relationship between the two motions can be an important factor.

Other terms that have been used to designate fretting-type damage are "false brinelling," which is particularly applied to bearings where the craters caused by the vibration of the ball against the race are circular and resemble Brinell impressions, and "fitting rust," where gages or shims are clamped together and experience vibration.

The terms "fretting" and "fretting corrosion" are now widely accepted in most languages, but the term "reiboxydation" is still used in Germany, and "corrosion de trepidation" is occasionally used in France.

Fretting was first reported by Eden, Rose, and Cunningham in 1911 (Ref 1), who found that brown oxide debris was formed in the steel grips of their fatigue machine in contact with a steel specimen. It was not until 1927 that Tomlinson (Ref 2) conducted the first investigation of the process and designed two machines to produce small-amplitude rotational movement between two annuli in the first case, and an annulus and a flat in the second case. The movement was controlled by a long lever system. Because the resultant debris on his steel specimens was the red iron oxide α-Fe_2O_3, which had arisen from chemical reaction with oxygen in the air, he coined the phrase "fretting corrosion." He also established that the damage could be caused by movements with amplitudes as small as a few millionths of an inch ($\sim$ 125 nm, or 1250 Å), and the important fact that relative movement had to occur, which he termed "slip."

The next important stage in the development of fretting studies was made by Warlow-Davies (Ref 3) in 1941, who produced fretting damage on the gage length of steel fatigue specimens and found a subsequent reduction in fatigue strength due to the pitting of the surface of between 13 and 17%. This was to be expected, but later investigations,

particularly by McDowell, Tierney, and McClellan (Ref 4) showed that the con-joint action of fretting and fatigue, which is the usual case in practice, was much more dangerous, producing strength reduction factors (SRFs) of 2 to 5 and even greater. Fenner and Field (Ref 5) in 1958 demonstrated that fretting greatly accelerated the crack initiation process. In normal fatigue, crack initiation may account for 90% of fatigue life, whereas in fretting fatigue, initiation could occur in $\leq$5% of the fatigue life.

Fretting Wear in Mechanical Components

Because vibration is one of the main causes of the fretting movement, it follows that the most likely area for it to occur is in machinery. The contacts between hubs, shrink- and press-fits, and bearing housings on loaded rotating shafts or axles are particularly prone to fretting damage, but because the movement arises from alternating stresses in the shaft surface, the problem is more one of fatigue than wear (Ref 6). However, wear rather than fatigue can be a problem in bearing housings. Thin-shell bearings are universally used in diesel engines, and such bearings involve an interference fit between the bearing and the housing. If the contact pressure is not high enough, movement can occur, giving rise to the fretting damage shown in Fig. 1 (Ref 7).

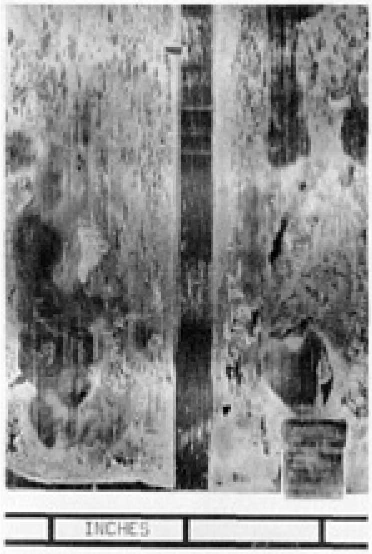


Fig. 1 Fretting damage incurred in a bearing shell component because of insufficient contact pressure between the shell and housing of a bearing. Source: Ref 7

Flexible couplings and splines, particularly where they form a connection between two shafts and are designed to accommodate some slight misalignment, can suffer severe fretting wear. Recent examples of such cases are described in the paper by Heinz (Ref 8).

Jointed structures are another source of fretting problems. According to Mitchell (Ref 9), there is no such thing as a static joint on an aircraft. An aircraft structure is largely an assemblage of aluminum and steel components riveted together or joined by fasteners. Even in a simple riveted joint, there are at least three possible sites where fretting can occur (Fig. 2):

- Between the riveted panels
- Between the underside of the rivet head and the panel
- Between the shank of the rivet and the rivet holes

Typical fretting damage is shown in Fig. 3.

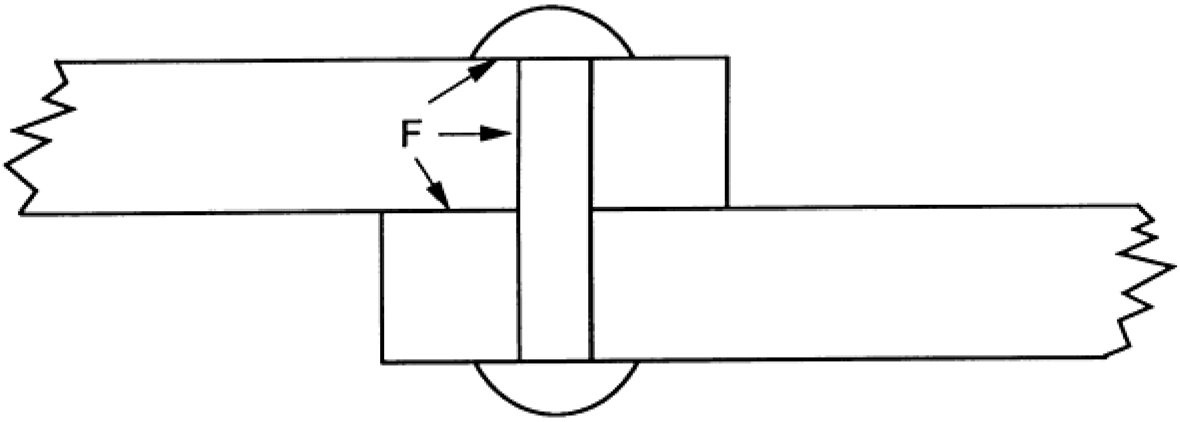


Fig. 2 Locations (indicated by F) prone to fretting in an aluminum-steel riveted joint

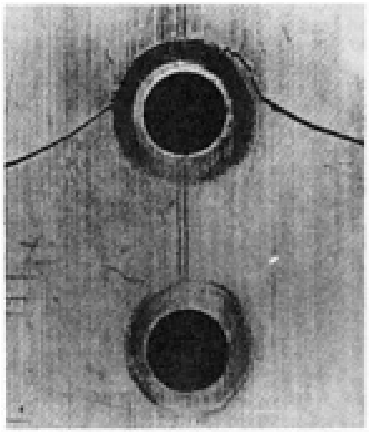


Fig. 3 Fatigue crack generated by fretting damage at a rivet hole

Turbines, both steam and gas, operate under conditions that subject these components to fretting damage. There are three main sites:

- Where the turbine disk is either shrink-fitted onto the driving shaft or attached to the shaft by means of a bolted flange
- Where the blades are fixed into the disk by either a dovetail or fir-tree type fixture
- Where snubbers at the outer ends of the blades contact those on adjacent blades

The first two sites are more likely to be areas for fatigue crack initiation, but loss of material by wear of the snubbers will give rise to increased vibration of the blades and the onset of impact fretting. In some turbines, a wire is threaded through holes in the outer ends of the blades to dampen down vibration. In this case, the frictional energy dissipated in the fretting process is the effective agent and may be said to be one of the few cases where fretting is beneficial.

The rotors of electrical generators have slots along their lengths in which the windings are located. These are held in place by wedges. Fretting can occur between the wedges and the undercut portion of the winding channel with the development of fatigue cracks in the rotor. These problems in turbines and generators are the subject of a recent review (Ref 10).

Steel ropes are used widely as mooring ropes, haulage ropes, mining ropes, and on cable cars. Overhead power lines are in effect ropes made up to aluminum wires wound on a steel support rope; more recently, they have been made up entirely of aluminum (the all-aluminum conductor). Any bending movement or alternating tension can result in local movement in

the numerous interwire contacts. Most ropes include some form of internal lubrication to lessen the effects of such movement. Again, one of the main dangers is fatigue failure of a wire, which increases the load on the remaining wires. However, in the locked coil rope that is designed with shaped wires to give an almost solid outer structure (Fig. 4), fretting debris with a greater volume than that of the metal from which it was formed can lead to the forcing apart of the wires, with the escape of lubricant and the ingress of a corrosive atmosphere. The accumulation of debris also leads to a reduction in flexibility at this point and, in the case of a mining rope, can cause jolting when this part of the cable passes on to the winding drum.

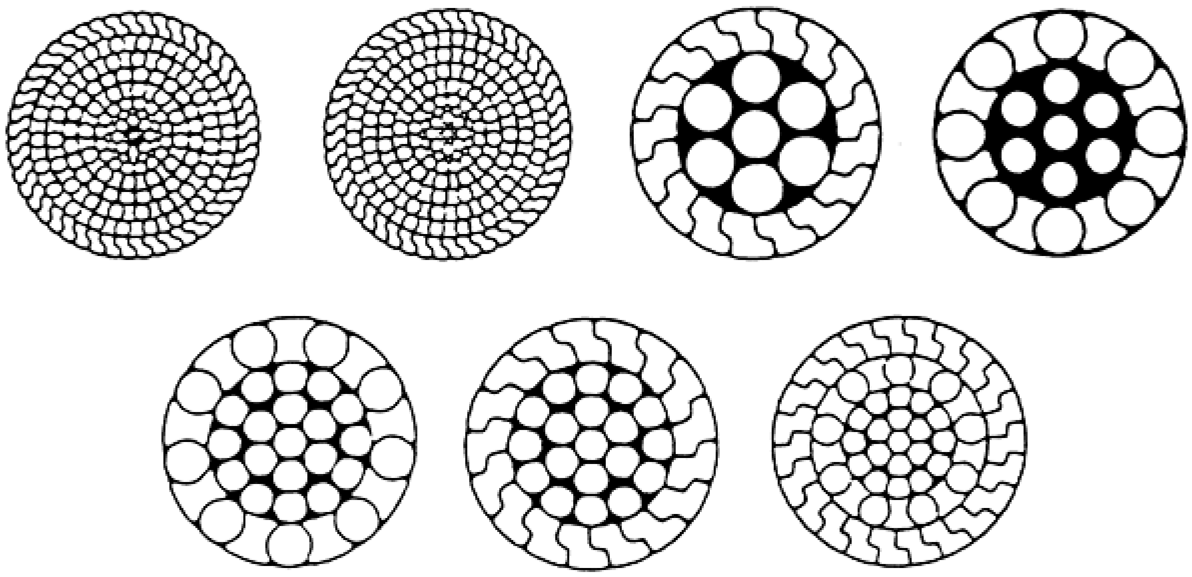


Fig. 4 Cross-sections of typical locked-coil track strand wire rope construction

In steam generators and heat exchangers, flow-induced vibration results in fretting between the tubes and the supports or baffles through which they pass. In the past, this appears to have been a particular problem in the nuclear power industry (Ref 11, 12) and chemical processing plants (Ref 13). A large contribution to the fretting in these industries is from impact fretting, which is discussed in the section "Impact Fretting" of this article.

Another example where movement is an integral part of the *modus operandi* of the device is the leaf spring used at one time in automobiles and railway rolling stock. It consists of an assemblage of steel plates, whose purpose is to dampen down vibration transmitted from the wheels to the carriage. Accumulation of debris could result in the plates not moving easily over one another and a stiffening of the spring.

The accumulation of debris is also the main problem in the fretting of electrical contacts (Ref 14). Because the oxide debris is nonconducting, interference or distortion of electrical signals can be generated in telephone equipment and also in computer connections. In some countries, aluminum conductors are used in domestic electrical circuits. Here the movement is due to temperature variations and is, therefore, of extremely low-cycle frequency. However, the buildup of aluminum oxide in connections can cause the formation of electrical arcs, which in extreme instances has led to houses being burned down.

The transportation of goods by road or rail obviously means that these are subjected to vibration. This gave rise to the classic case of fretting in the bearings of automobiles being transported by rail from Detroit to the West Coast of the United States. As the bearings were stationary, fretting occurred between the loaded balls and the race, giving rise to the false brinelling mentioned above. This has also been a problem with the transport of stacks of aluminum sheets with highly polished finishes. It also led to the abandonment of using stacked aluminum trays to transport food stuffs. Another example involved the transport of live ammunition in aluminum boxes, because the fretting debris was found to be pyrophoric. In a case investigated by the author, high-purity (HP) aluminum "logs" (cylinders 250 mm, or 10 in., in diameter), were transported by rail to the plant where they were to be extruded as buss bars. Black marks (that is, fretting debris) were noticed on the cylinders (Fig. 5) that gave rise to serious surface tears on extrusion (Fig. 6).

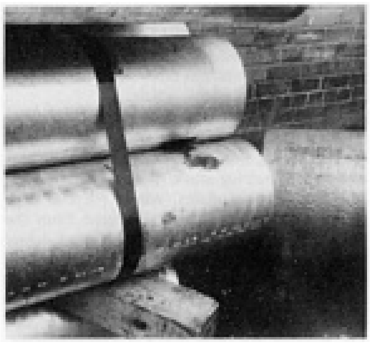


Fig. 5 Typical fretting marks on a high-purity aluminum cylinder

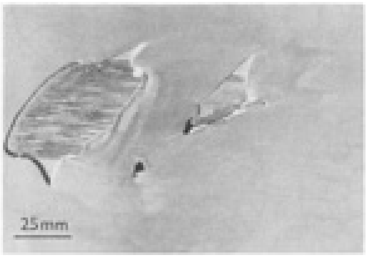


Fig. 6 Surface tears caused by fretting damage when aluminum cylinders in Fig. 5 are extruded

Fretting can also occur in certain orthopedic devices, particularly fracture fixation devices such as bone plates. These are used to hold together the parts of fractured bone. They consist of plates of corrosion-resistant materials that are attached to the bone by means of screws of the same material. Fretting can occur between the underside of the screw head and the countersink of the hole in the plate. Fretting in this case results in disruption of otherwise protective films and the continual release of heavy metal ions into the body fluids, with possible toxic reactions. This aspect of fretting in aqueous electrolytes is dealt with in detail below. Fatigue failure is also possible (Fig. 7).

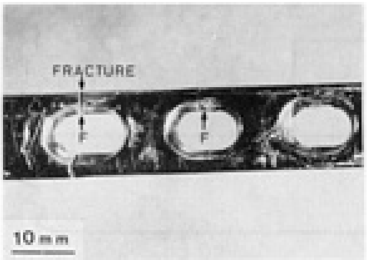


Fig. 7 Fracture in an orthopedic bone plate. A failure was caused by fretting damage (loss of protective oxide layer) in the countersunk portion of the plate.

The possible situations in which fretting can occur are legion, and it is only possible in this review to give some typical examples. In most of the cases described, suitable methods of overcoming the problem have been forthcoming as the result of detailed investigation; these are described in detail in the section "Prevention of Fretting Damage" in this article.

A few years ago, the author was approached by a professor of dentistry from Paris, who was convinced that the spots of damage (caries) that develop between contacting teeth, particularly the molars, were a form of fretting corrosion due to the relative movement occurring during chewing and swallowing, and certainly the examples he produced looked very much like it. This extends the possibility of fretting damage to materials other than metals. With increased use of

polymers, both simple and reinforced, in addition to ceramics, this is an area where further examples of a somewhat different kind are likely to be encountered. Some research already is being devoted to the topic and is described in the section "Parameters Affecting Fretting" in this article.

Parameters Affecting Fretting

Experimental investigations have concentrated on the effect of specific physical variables (such as amplitude of slip, normal load, frequency of vibration, and the circumstances of the fretting situation, such as type of contact, mode of vibration, and the condition of the surfaces) to indicate how the problem can be overcome in the future design of contacting components. These results are discussed in detail below.

Amplitude of Slip. Tomlinson (Ref 2) established that relative movement was essential for fretting to occur and showed that extremely small movements of the order of a few nanometers were capable of causing damage. This aspect of fretting--the lower limit of slip--has more recently been the subject of investigation (Ref 15, 16). One of the problems in experimental work has been the control and measurement of small amplitudes of movement. The most satisfactory method which avoids possible losses due to elastic deformation in the supports or actuators, is to concentrate on the situation of partial slip. This is best studied with a contact consisting of a ball (typically a bearing ball) on a flat, a situation that was analyzed by Mindlin. Figure 8 shows the form of contact of a ball on a flat developed by Mindlin from the Hertzian solution for a stationary contact. On the application of a tangential force or a torsional force, slip occurs over an outer annular region of the circle of contact. If the applied force is oscillating, then fretting occurs in the slip region. Johnson used this method to verify the Mindlin analysis (Ref 17). More recently, Nowell and Hills (Ref 18) have shown that the stress distributions are somewhat distorted but the basic idea remains the same. The center of the contact region remains locked (that is, the nonslip region). The amplitude of movement in the slip region for torsional excitation presumably rises from zero at the inner boundary to some greater value at the outer edge. The dimensions of the damage area closely approach these limits.

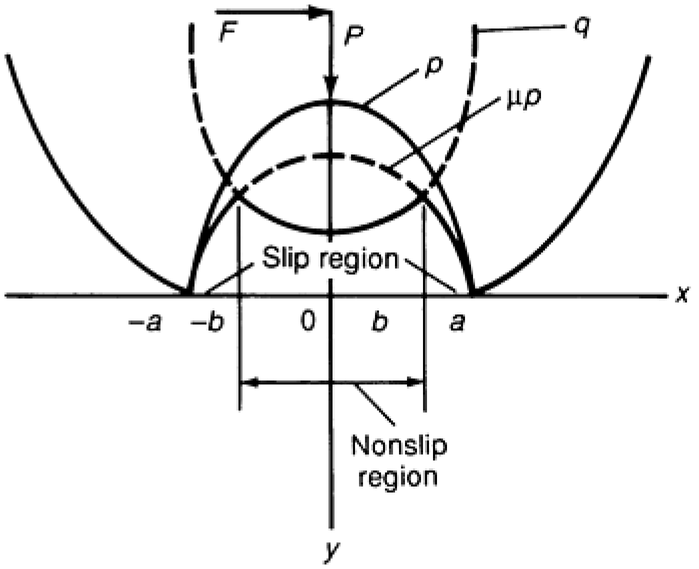


Fig. 8 Mindlin analysis of partial slip for a ball in contact with a flat and subjected to a tangential force

In the investigations by Kennedy and his colleagues, the amplitude of slip at which damage was detected on steel specimens under a load of 88 N (8.8 kgf) was 0.06 μm (2.4 μin.). A narrow annulus of damage was detected at an amplitude of 0.1 μm (4 μin.). As the amplitude was increased, the damage area, characterized by the appearance of the typical reddish-brown oxide, increased in width, with the outer radius remaining constant. This situation has been investigated by Kuno (Ref 19) using a linear vibratory motion, where it was found that the annulus of damage for a fixed amplitude increased in width with the number of fretting cycles, but the outer radius remained constant (Fig. 9).

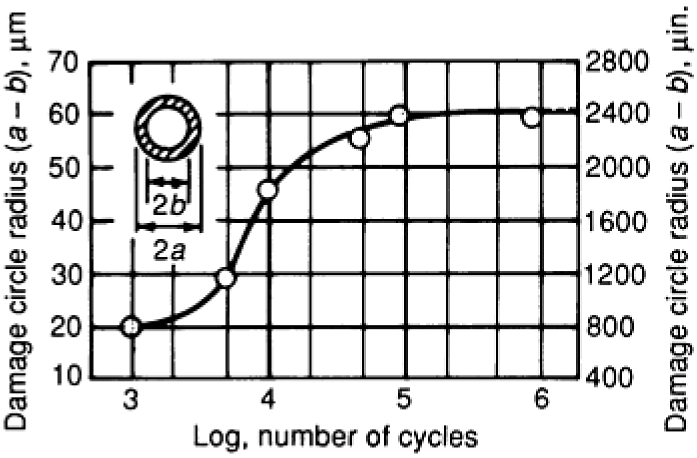


Fig. 9 Relation between diameter of fretting scar and the number of cycles for a ball on a flat under conditions of total slip. *P,* 437 N (43.7 kgf); P_{max}, 1.2 GPa (175 ksi); amplitude of movement, 8 μm (320 μin.); frequency, 50 Hz

In the Kennedy experiments, the damage was detectable by profilometry of the flat surface at an amplitude of 0.53 μm (21 μin.) and was said to be severe at 2.8 μm (110 μin.). The level of slip amplitude at which slip occurs over all of the contact area depends on such factors are the normal load and the mechanical properties of the materials. In the heavy loads used by Goto (Ref 20), which were sufficient to produce a Brinell impression, the slip annulus was extremely narrow and it is doubtful whether total slip could ever be achieved under these conditions (Fig. 10). Under more normal conditions (amplitude 8 to 25 μm, or 320 to 1000 μin.; normal load 135 N, or 13.8 kfg) Kuno found that total slip occured at an amplitude of 25 μm (1000 μin.). At this amplitude and above, the radius of the damage circle increased linearly with the number of cycles.

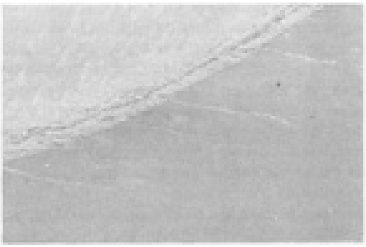


Fig. 10 Annulus of fretting damage produced by contact of a ball on a flat under heavy normal load

The amplitude of slip where the slip becomes total has been termed the critical amplitude, Δ. Vingsbo and Söderberg (Ref 21), using a crossed-cylinder arrangement with steel specimens, have mapped the relation between the critical amplitude and normal load (Fig. 11). These results are somewhat different from those of Kennedy, whose experimental values agree well with calculations based on elastic theory (Fig. 12). The discrepancy is attributed to plastic deformation occurring in the crossed-cylinder experiments (Ref 22).

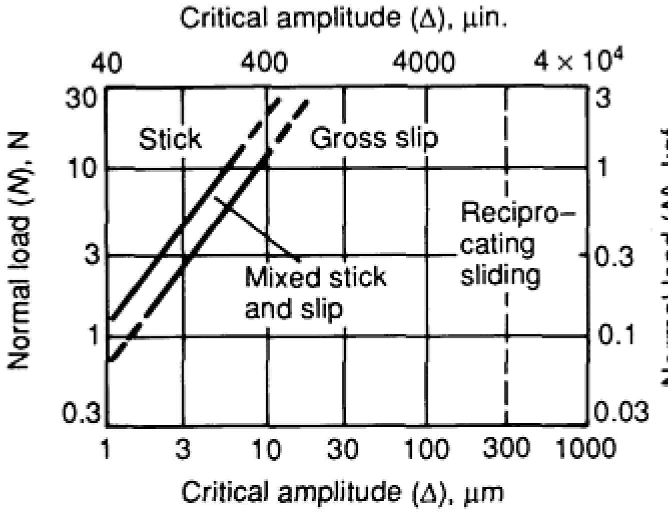


Fig. 11 Plot of normal load versus critical amplitude as a function of normal slip for a crossed steel cylinder arrangement. Source: Ref 21

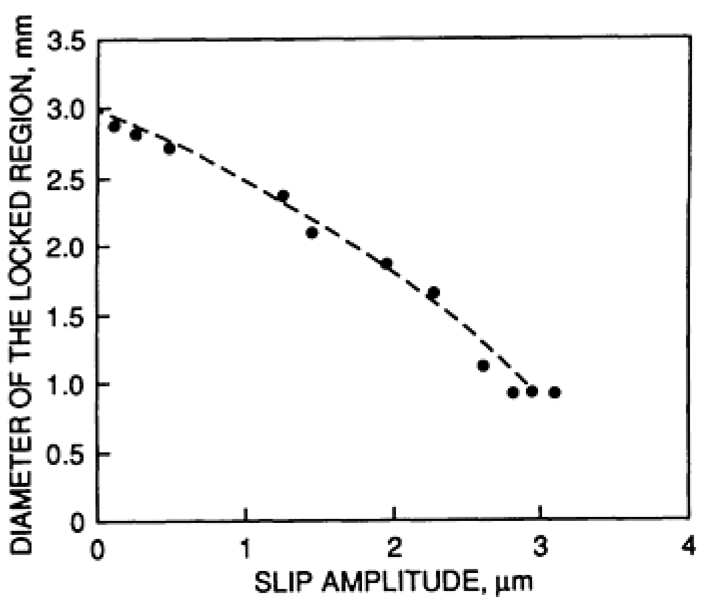


Fig. 12 Comparison of the theoretically predicted diameter of the nonslip region versus slip amplitude based on experimental data

In other investigations, using a variety of different contact geometries and different methods of assessing the wear rate and plotting the results as the specific wear rate (that is, the loss of material per unit of sliding distance per unit of applied normal load), the curves are of the form shown in Fig. 13. The curves are sigmoidal, with the specific wear rate becoming constant at amplitudes >100 μm (>0.004 in.) (that is, identical with unidirectional sliding or reciprocating sliding wear rates). This then gives a possible upper limit for the slip amplitude for the case of true fretting.

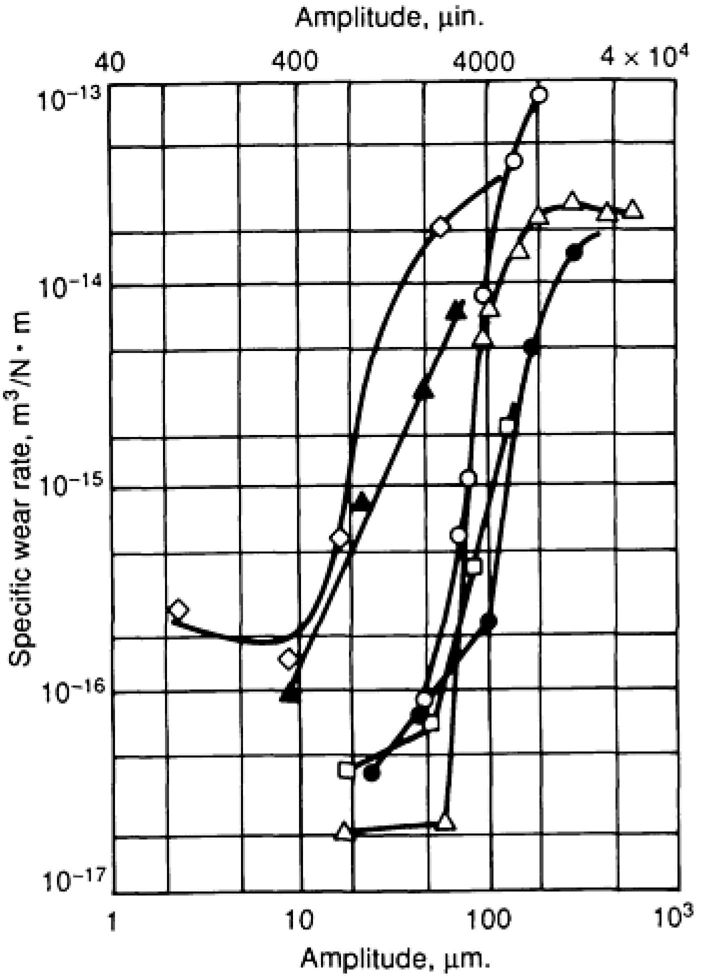


Fig. 13 Plot of specific wear rate versus amplitude of slip. Each core is the result of a separate investigation.

The small-amplitude characteristic of fretting is the basis of many of the important features of the process. It means that the relative velocities of the two surfaces are much lower, even at high frequencies, compared with conditions in unidirectional sliding. At a frequency of 25 Hz and an amplitude of 25 μm (1000 μin.), the average velocity is 1.25 mm/s (0.0492 in./s), whereas sliding speeds in bearings are of the order of 1 m · s. Debris tends to stay where it is generated, so that high spots (for example, machining marks) become imprinted on the other surface if it has a higher degree of finish (Fig. 14). The debris may never become exposed to the atmosphere and can become compacted under the compressive load, often changing its appearance, as discussed in the section "Effect of Debris on Wear" in this article.

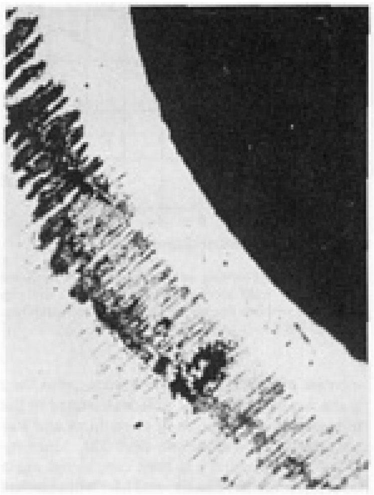


Fig. 14 Marks produced by a machined surface fretting against a polished flat

Normal Load. When two surfaces are placed in contact and loaded, plastic deformation of contacting high spots (asperities) occurs, as established by Bowden and Tabor (Ref 23). However, certain contacts (for example, sphere on flat and cylinder on flat), where the contact area tends to be extremely small, still conform to the elastic theory of Hertz because the overall contact is determined by the elastic stress field developed in the hinterland or bulk of the material. In most practical cases of contact between conforming surfaces, plastic deformation is to be expected at containing high points. The real area of contact, *A,* is therefore directly proportional to the applied load, *P:*

$$A = \frac{P}{p_0} \quad \text{(Eq 1)}$$

where p_0 is the yield pressure and roughly equal to 3*Y,* and where *Y* is the yield stress in tension. Because fretting only occurs at contacting areas, it follows that the amount of wear should be directly proportional to the applied load. This is generally found to be the case if the amplitude of slip is kept constant. In the early stages there is a running-in period in which metal-to-metal contact predominates (Fig. 15). This can be detected by measurement of contact resistance which remains low and depends for a given number of cycles on the magnitude of the load (Fig. 16). These observations are discussed in greater detail in the section "Mechanism of Fretting Wear" in this article.

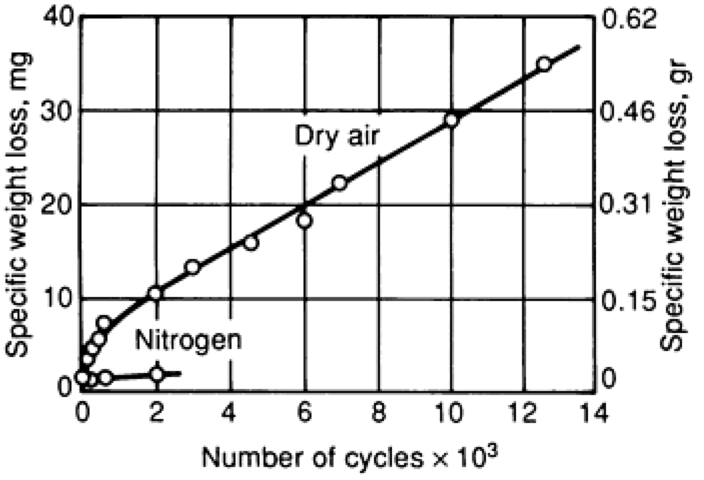


Fig. 15 Plot of fretting wear versus number of cycles for mild steel with 90 μm (0.0036 in.) slip amplitude in

both dry air and nitrogen atmospheres. Source: Ref 24

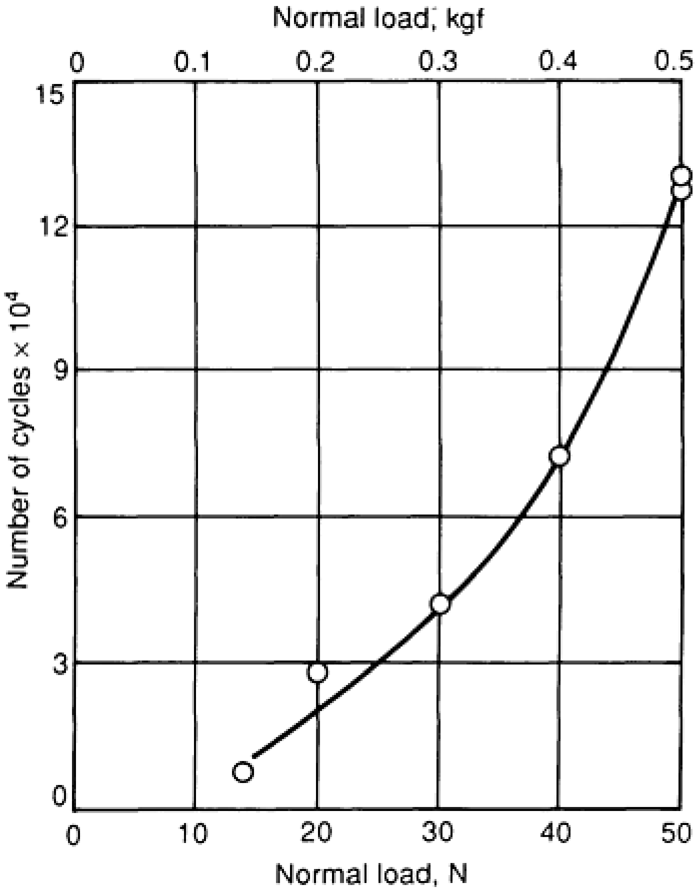


Fig. 16 Relation between initial period of metal-to-metal contact and normal load

Frequency. The effect of frequency was originally studied by Uhlig (Ref 24), who found an increase in wear volume at low frequencies for a given number of cycles. This was related to the formation and disruption of oxide films and was considered by Waterhouse (Ref 25). More recently, the question has been considered again by Söderberg, Bryggman, and McCullough (Ref 26), who were looking at a means of speeding up fretting wear tests, assuming that high frequencies had no significant effect. The range of frequencies investigated was from 10 to 10^4 Hz, using steel specimens in a crossed-cylinder contact. Their conclusion was that, in the partial-slip situation, although there was little effect on measurable wear, the increase in the interfacial strain rate at high frequencies led to increased fatigue damage and increased corrosion due to the rise in temperature. However, in the total-slip situation, there was little effect over the frequency range 100 to 20×10^4, and these high frequencies could be used in accelerated tests.

Not all movements that can lead to fretting damage take place at anything like the frequencies mentioned above. Relative movement can be due to differential thermal expansion, which, if arising from diurnal changes in temperature, can have a frequency of only a few cycles a day. The pressurizing of airplane cabins is in a similar frequency range. The movement of mooring cables due to water flow is of considerably higher frequency but is assumed to be ~0.1 Hz. In such circumstances, the environmental and corrosion effects have much greater importance, as can be seen by the extrapolation of Uhlig's results to lower frequencies (Fig. 17).

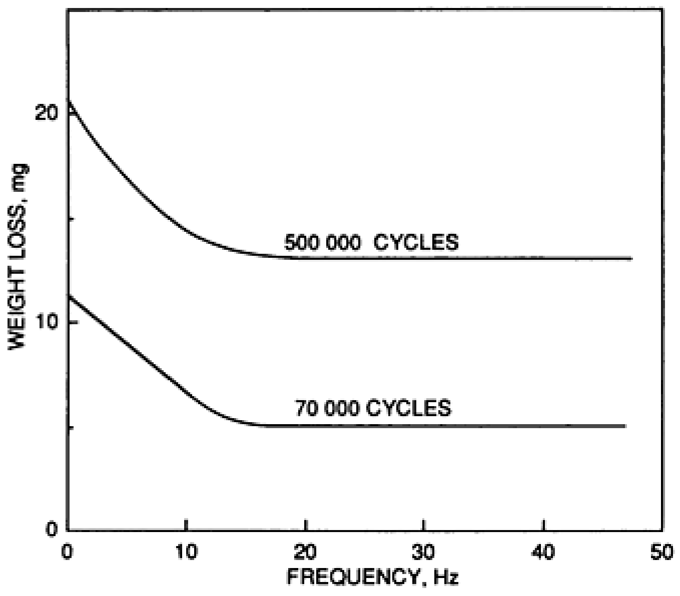


Fig. 17 Plot of wear versus frequency of fretting vibration

The area of low frequency has been mainly investigated in relation to electrical contacts. In many of these investigations, the low-frequency movement (0.003 to 0.03 Hz) has been produced by cyclic heating and cooling of a copper rod. This allows also very low amplitudes to be investigated (Ref 27). The extensive work of Antler (Ref 28) in the United States and of Braunovic (Ref 29) in Canada has explored the effect of different contacting surfaces on the contact resistance under fretting conditions. In these studies the aim has been to maintain metal-to-metal contact throughout the working life of the contact. A complication in these experiments is the formation of a friction polymer from carbonaceous impurities in the atmosphere, which can have a profound effect on the contact resistance. For separable connectors, gold plating appears to be the most satisfactory surface, although cheaper coatings are actively being sought (Ref 30) and are discussed further in the section "Prevention of Fretting Damage" in this article.

Type of Contact. As will be seen later when the role of fretting debris in the process is discussed (see the section "Effect of Debris on Wear" in this article), the ease with which debris can escape from the contact region is an important factor in the fretting process itself. It also gives rise to further problems by contamination of other surfaces (in a machine, for example).

Tomlinson's original design of an apparatus with torsional vibration of annulus on flat has much to commend it because no part of the contacting surfaces becomes exposed, and, for debris to escape, it must move at right angles to the direction of motion. A further slight advantage is that the amplitude of slip has a small variation from the inner to the outer edge and can therefore be used to investigate the effect of amplitude in one test. With flat contact surfaces, however, the initiation and development of areas of wear damage are sporadic no matter how carefully the surfaces are prepared and the alignment controlled. Contact pressure is usually expressed as the nominal value calculated from the apparent area of contact and the applied load. Measurement of the real area of contact would require contact resistance measurements at the outset of the test, or it could be calculated with a knowledge of the yield pressure (Eq 1). Contacts consisting of geometrical surfaces such as sphere on flat, cylinder on flat, and crossed cylinders allow some initial calculations based on elastic theory of the area of contact and stress distribution therein. However, these quantities are rapidly changed as the wear process proceeds. Also, such contacts increase the possibility for debris to escape. It has recently been shown that the escape of debris in the crossed-cylinder arrangement is greatly influenced by the direction of motion (Ref 31). The arrangement shown in Fig. 18 (b) allows the debris to escape by being pushed out by the axial movement of the upper cylinder, leading to more frequent metal-to-metal contact and a higher wear rate than the arrangement shown in Fig. 18 (a). However, in experimental investigations, the ease and cost of preparing specimens are significant factors, and the crossed-cylinder arrangement is one of the most convenient. It also mirrors the situation of contact between wires in steel ropes.

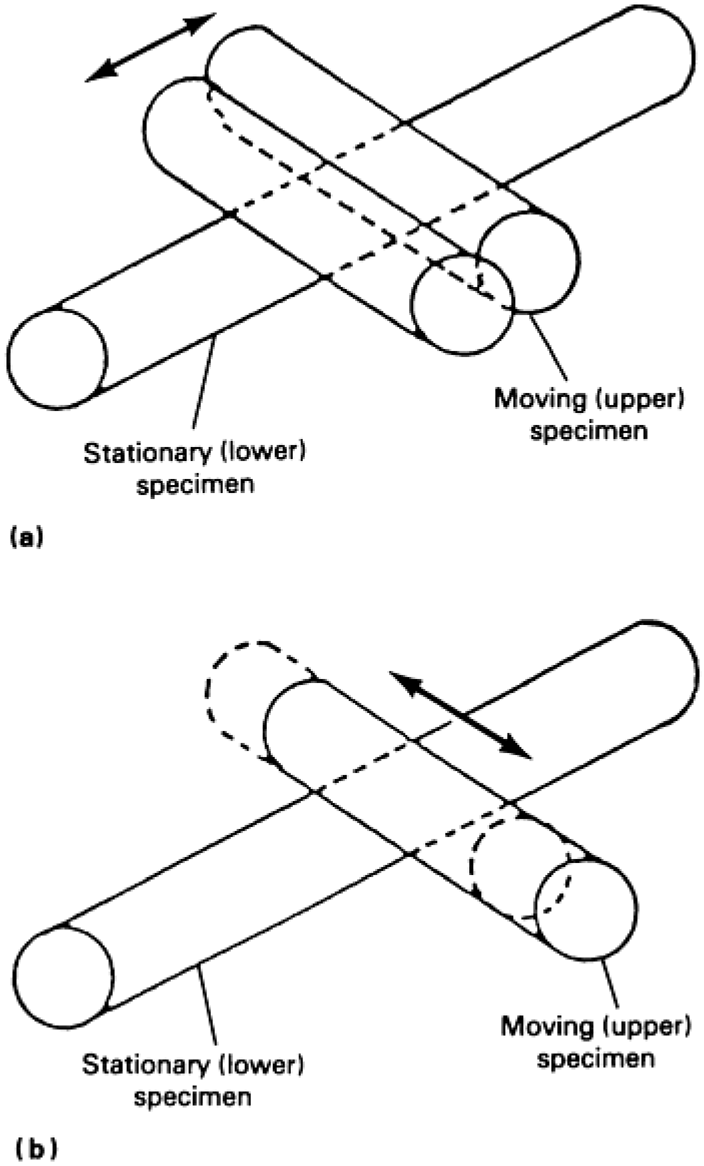


Fig. 18 Two test directions for determining fretting in a crossed-cylinder arrangement. (a) Parallel to axis of lower specimen. (b) At right angles to axis of lower specimen

Type of Vibration. In most experimental investigations, the vibration is simple harmonic motion of constant amplitude. This may be the case in practical cases of rotating shafts, but is very far from situations in which frequencies may vary and pass through areas of resonance. In addition, aircraft structures will be entirely under the influence of random loading with no fixed frequency or amplitude. The same can be said of ropes and cables in moorings, overhead ropeways, and elevator ropes.

There have been few investigations into this aspect of fretting wear. Elliott, Mabie, Furey, and Mitchell (Ref 32) conducted a detailed analysis of the vibrations in one of the bearings supporting the shaft of an ac/dc generator operating onboard a ship. Spectra of the relative displacements in the axial, tangential, and radial directions were determined in the frequency ranges 0 to 500 Hz and 0.5 to 10 kHz, with no load and under full load. Although it was not possible to distinguish between elastic movement and rigid body movement, it was assumed that amplitudes >1 μm (>40 μin.) were due to the latter. The amplitude range was from 4.30 to 10^{-4} μm (173 to 0.004 μin.) in the frequency range <250 Hz, with several of the peaks being >0.05 μm (>2 μin.). The tangential component under full load had an upper limit of 8 μm (320 μin.) and combining this with the movements in the axial and radial directions gave a vector sum of 120 μm (480 μin.), at which considerable fretting damage was experienced. Their analysis indicated that there was significant relative motion normal to the fretting surface.

This latter observation has been investigated by Golego and his co-workers (Ref 33). The construction of their tribometer allowed frequency variations, phase differences between tangential and normal vibration, and the control of resonance. They identified three areas:

- Subresonant loading conditions in which the damage was said to resemble typical fretting damage
- Resonant conditions in which the surface damage was significantly less due to the lower dissipation of energy due to friction
- Superresonant operation, which showed tearing of the surface that was associated with a phase shift of 90° between the tangential and normal vibrations

Impact Fretting. In the section "Type of Vibration," the effect of a cyclically varying normal force superimposed on the tangential movement was considered for situations in which the two surfaces remained in constant contact. The circumstances where the two surfaces can become separated during some part of each cycle are somewhat different, because of the contribution of impact. As mentioned earlier in the section "Fretting Wear in Mechanical Components," this is a problem frequently encountered in heat exchangers, where pipes or fuel rod containers are restrained by supports or baffles but experience vibration as a result of fluid flow.

Two types of wear can be distinguished in the practical case encountered in nuclear reactors. The first involves whirling of the tube resulting in intermittent unidirectional sliding and is obviously not fretting. The small-amplitude movement that results in fretting-type damage is the result of flow-induced vibrations. The important parameters are impact force, clearance between tube and support, support plate thickness, and frequency. The wear rate for stainless steel increases by an order of magnitude for a tenfold increase in impact force, and a similar increase for a fourfold increase in the clearance. Increasing the support plate thickness from 10 to 20 mm (0.4 to 0.8 in.) reduces the wear rate from 0.7 to 0.2 μm (30 to 8 μin.) per 10^6 cycles. At frequencies below 40 Hz, there is no effect of frequency, but at higher frequencies the wear rate is reduced due to the reduction in contact time per cycle (Ref 34). These results were obtained at room temperature, whereas in operation these devices would be at a higher temperature and thus the wear rates could be profoundly influenced both by temperature and environment; such effects are discussed in the section "Environmental Effects."

Many years of work by de Gee (Ref 35) indicated that the combination of impact and fretting was more severe than fretting alone, whereas much of the work on tube and tube support vibration suggests that the impact contributes little to the wear rate and that continuous contact is more damaging (Ref 36, 37). An important variable in considering this type of motion is the phase relation between the normal and tangential vibrations. Differences are to be expected between impact at maximum tangential velocity and zero velocity. Also, the time period per cycle that contact is made will be of great significance. In Ref 34, the following contact times per cycle are quoted as a function of frequency:

Frequency, Hz	Contact time per cycle, s	Difference between impact at maximum tangential velocity and zero velocity, %
26	0.022	57
40	0.010	40
60	0.005	30

Rigs have been constructed to study the effects of these phase relationships, particularly under severe environmental conditions, and those results that have been published are referred to in Ref 34.

Surface Finish. Early observations of practical examples of fretting damage suggested that the more highly polished the contacting surfaces, especially if flat, the worse the damage. On polished aluminum where the debris is largely Al_2O_3 (corundum) with a high hardness, it could spread the damage, whereas on a rough surface the debris can escape into the hollows between the real contact areas. There is also the possibility that on a rough surface, if the exciting force is of small amplitude, some of the relative movement will be taken up by elastic deformation of the asperities. Some work by Calhoun shows that a rougher surface suffers less damage (Ref 38).

Shot peening is a widely applied surface treatment that produces roughening, work hardening, and a residual compressive stress in the surface. The author showed that shot peening a steel surface reduced the coefficient of friction in fretting (Ref 39), but this could also be due to work hardening and possibly also to the compressive stress. To investigate this phenomenon further, and to try and separate the three effects of shot peening, a series of tests was carried out on treated cylindrical steel specimens. The residual stress and work hardening could be removed by annealing, the residual stress could be reduced by controlled tensile straining, and the surface roughness could be removed by careful surface polishing (Ref 40). The shot-peened surface of 0.4% C steel had a very low value for fretting wear; however, polishing raised the level of wear to that of the unpeened steel.

Residual Stresses. Because one of the possible mechanisms of fretting damage is surface fatigue, it is to be expected that a residual stress field will have some effect on the wear process. Experiments were performed with a bearing ball on a bearing steel (1.0C-1.5Cr) specimen that was statically bent to give residual stresses over a range -900 to 900 MPa (-130 to 130 ksi). When the specimen was vibrated with a contact load of 216 N (22 kgf) and an alternating tangential load of ±20 N (±2 kgf) so that partial slip was occasioned, the depth of the wear scar increased with tensile stress but decreased with compressive stress (Fig. 19).

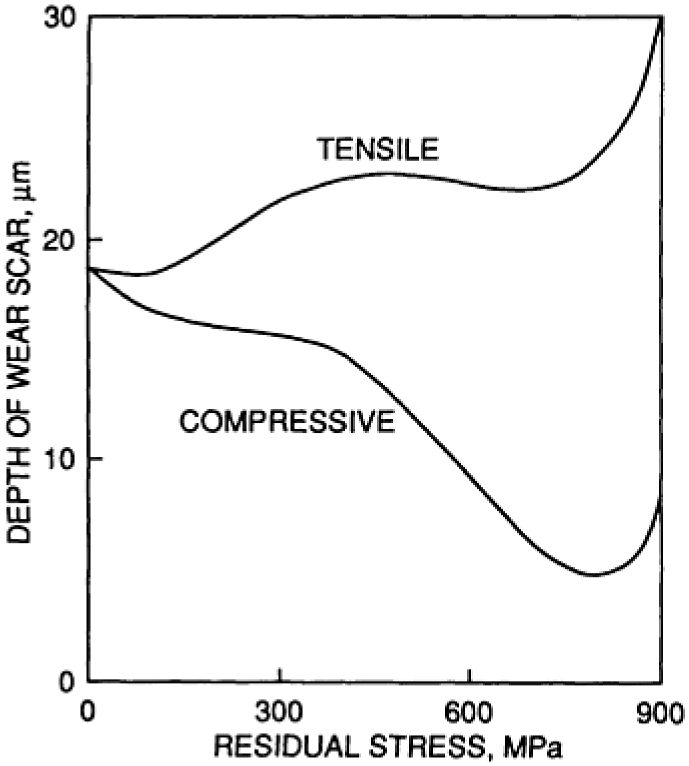


Fig. 19 Plot of depth of wear scar versus compressive and tensile residual stresses in the fretted surface. Source: Ref 41

Endo had shown earlier that residual stresses can be generated by the fretting action (Ref 42) and this has been confirmed in further work by Yournadj, Maeder, and Moisan (Ref 43). In fretting, flat steel surfaces with residual compressive stresses up to 735 MPa (107 ksi) were generated; these were most important in a direction at right angles to the direction of motion. The converse of this effect, namely the modification of existing compressive stresses by fretting, has been observed by Fair, Noble, and Waterhouse (Ref 44) in fretting-fatigue experiments. The stresses fade under normal fatigue but disappear much more rapidly under fretting fatigue (Fig. 20).

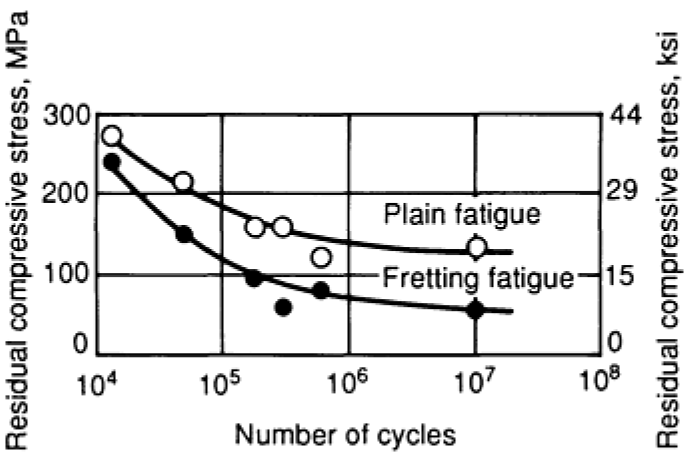


Fig. 20 Fading of surface compressive stress induced with a number of fretting cycles by shot peening

Finally, the presence of a static tensile stress can lead to the propagation of cracks initiated by the fretting (Ref 45).

Materials

If fretting wear is a possible occurrence in the design of a machine or structure, it may be possible to reduce the effect by choosing materials that are more resistant to such damage. This section deals with bulk materials not coated or surface treated; coated and surface-treated materials are dealt with in the section "Prevention of Fretting Damage" in this article. In addition, information on the behavior of new materials such as composites and ceramics must also be considered because of their increasing use in engineering applications.

Steels have been most extensively studied, largely because of their ability to change mechanical properties over a wide range by heat treatment without changing their chemical composition. In experiments with a steel ball on a steel flat, in which the hardness was varied from 220 to 850 HV, there was little difference in the wear loss (Ref 46). In crossed-cylinder experiments with three alloys of the same crystal structure, increasing hardness is shown to lead to a decrease in the critical amplitude between partial and total slip (Ref 47). The materials were copper, copper-silicon, and austenitic stainless steel, all materials of low stacking fault energy; as a consequence, work hardening occurred at low frequencies but softening set in at high frequencies because of the temperature rise.

In some investigations, notably those by Dobromirski (Ref 48) and Beard (Ref 49), the formation of a white etch-resistant layer has been detected in fretting of carbon steels (Fig. 21). This material is extremely hard and tends to develop cracks. It is often referred to as martensite, and some authors have regarded this as evidence that high local temperatures can be developed in fretting. The question of temperature rise is controversial and is discussed more fully below. However, if there is a temperature rise, materials that rely on low-temperature heat treatment for their optimum mechanical properties (for example, most aluminum alloys) are likely to suffer a further type of damage in fretting.

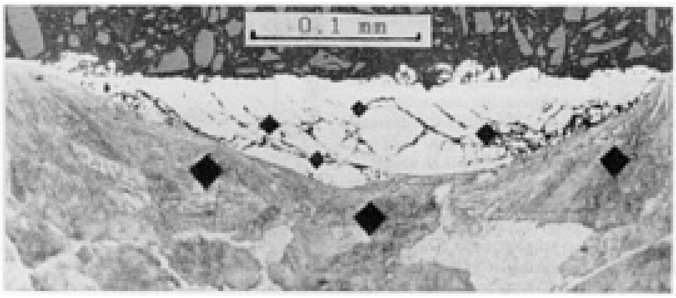


Fig. 21 Cross-sectional view of a white layer of martensite produced by fretting of a carbon steel connecting rod. Axial stress, 0 to 380 MPa (0 to 55 ksi); contact stress, 40 MPa (6 ksi); fretting cycles, 10^5. Sample was

nital etched and viewed with scanning electron microscopy (SEM).

A recent development has been the rapid quenching of liquid metals to produce amorphous metals. Fretting of two of these materials, Co-4Fe-2Mo-16Si-12B and Fe-4Si-15B, indicates that they are not wear resistant and become embrittled by the fretting action (Ref 50).

Metals and alloys that rely on a protective oxide film for corrosion resistance can suffer considerable fretting damage in environments (for example, ambient air), that would not be considered particularly corrosive. These include the more reactive metals such as aluminum, titanium, zirconium, and chromium when used as elements to alloy stainless steels. This aspect of fretting is considered in the section "Environmental Effects" in this article.

When dissimilar metals pairs are fretted against each other, their mutual solubility is a factor in the severity of the damage. Nickel, chromium, and iron are severely damaged when fretted against iron or chromium because of high mutual solubility; however, the damage is mild when these metals are fretted against copper or silver because there is no solubility (Ref 51).

In the fretting of polymers against steel, the possibility exists that one material will be transferred to the opposite surface. Damage arises in this case from platelike iron oxide particles adhering to the polymer; thus, the process becomes the fretting of two oxide-coated surfaces. Damage can occur to the polymer by the formation of cylindrical wear particles formed by rolling up of thin detached surface layers, which are extruded in a direction at right angles to the motion (Ref 52, 53).

In an investigation of fretting wear of fiber-reinforced polymers (composites) in contact with an aluminum alloy, weight loss was used to measure the damage (Ref 54). At amplitudes <900 μm (<0.036 in.), the wear was low and showed a slight increase with increasing amplitude. Above 900 μm (0.036 in.), the wear accelerated. Wear rate showed a stepwise increase with frequency, the critical frequency being 35 Hz at room temperature but reduced to 20 Hz at 80 to 100 °C (175 to 212 °F) due to a local rise in temperature. Unlike metals, which show a linear increase in wear with increasing normal load, these materials show little effect of load up to a critical value at a pressure of 20 MPa (3 ksi), above which a rapid increase occurs. Fiber orientation with respect to the sliding direction is another variable with these materials. Here the specific wear rate was negligible at angles of 0 and 90° between fiber axis and sliding direction, but showed a pronounced peak at 45°, as did the coefficient of friction (Fig. 22).

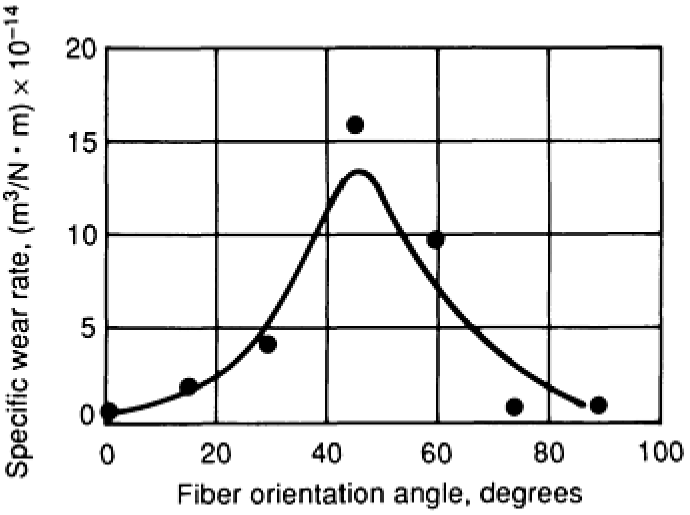


Fig. 22 Plot of fretting wear versus fiber orientation angle in a composite

In experiments involving the fretting of sintered alumina against a wide variety of metallic materials in a crossed-cylinder configuration, no detectable wear was found on the alumina, although metal transfer to its surface occurred (Ref 55). A similar result was found by Klaffke in fretting a steel ball against alumina, silicon nitride, and zirconia in dry air; however, silicon carbide produced equal wear on the ceramic and steel (Ref 56). An increase in relative humidity

generally resulted in a decrease in wear rate, except in the case of silicon nitride, which exhibited behavior similar to that silicon carbide.

It appears that when oxides such as alumina and zirconia are fretted, there is little damage. Fretting alumina against alumina produced no measurable wear, although some polishing of the surface was apparent (Ref 31, 57) (Fig. 23).

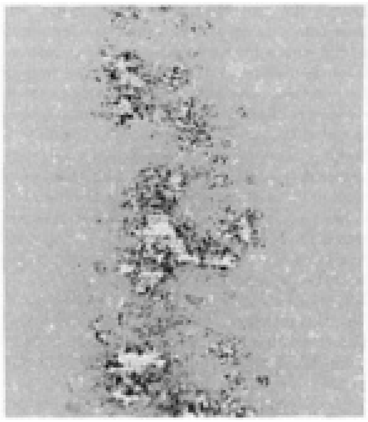


Fig. 23 Fretting mark generated by fretting an alumina cylinder against an alumina flat

In one of the earliest observations of fretting wear, Godfrey fretted a steel ball against the underside of a glass microscope slide and observed the process in the microscope (Ref 58). This type of experiment has been repeated in Japan, but no quantitative measurements of wear have been made. However, valuable information has been gained on the initiation of cracks in the glass, a typical brittle material, and on the onset of damage to the steel, and is therefore relevant to discussions on fretting mechanisms (Ref 59).

Effect of Debris on Wear

Debris oozing from a joint or connection in machinery is often the first indication that fretting damage is occurring. In recent years, the role of the debris in wear processes has been given considerable attention because it can build up between the sliding surfaces and increase the number of interfaces from one to two (Ref 60). This is particularly the case in fretting where escape of the debris is greatly restricted compared with unidirectional sliding. This section will focus on the form, composition, and role of the debris.

It is well established that the debris on mild steel and other steels in air is essentially α-Fe_2O_3, which when finely divided, is reddish brown in color and nonmagnetic. Its hexagonal crystal structure is identical with that of α-Al_2O_3. Under high compressive loads, it can become compacted, and its color is then black and iridescent. In this respect, it is identical to the mineral hematite, which in massive form is blackish but is red when powdered. On aluminum, the debris is black and consists of aluminum particles surrounded by Al_2O_3. The aluminum content is 23% and the powder is pyrophoric (Ref 61). In other metals, such as copper and titanium, the debris is the expected oxide. In some cases, depending on the method of its production, the debris can contain some unoxidized metal.

The debris is usually in the form of thin platelets a few microns thick and 50 to 100 μm (0.002 to 0.004 in.) in diameter, but it can be much smaller, as in Fig. 24, where the debris resulted from fretting steel against a polymer. The particle size was shown in this case to be related to the surface energy of the polymer. Fretting sometimes occurs between the walls of fatigue cracks, when spherical and cylindrical shapes of debris may result due to the very small movements and the high compressive stresses during the compression part of the fatigue cycle.

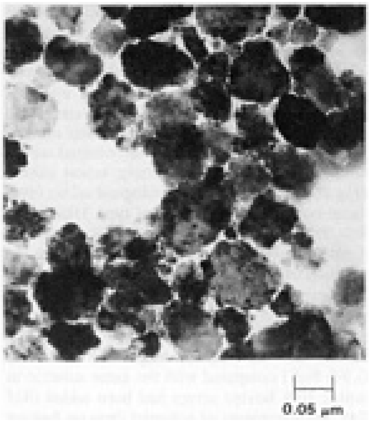


Fig. 24 Oxide debris produced when fretting steel against a polymer

The role of the debris on steel specimens has been demonstrated by carrying out tests where the fretting surfaces were periodically separated and the debris removed (Ref 62), resulting in an increase in the wear rate. Conversely, the introduction of α-Fe_2O_3 powder between the surfaces reduced the wear rate; the resultant surfaces were hardly damaged and showed signs of slight polishing. In long-term experiments, the terminal stage of fretting was identified as the formation of extensive beds of compacted oxide with a much reduced wear rate (Ref 63). With very small slip amplitudes of less than 10 μm (400 μin.), the wear ceases because the oxide beds support the load and separate the surfaces (Ref 64). At higher amplitudes of $\geq$20 μm ($\geq$800 μin.), the wear rate is much reduced; however, wear continues, and occasional metal-to-metal wear can still be detected by contact resistance measurements, even after millions of cycles (Ref 65). These observations have considerable bearing on the ideas that have been put forward for explaining the mechanism of fretting, which are discussed below. The conclusion is, however, that the accumulation of debris between the surfaces in compacted layers is more likely to reduce further wear and that abrasion by the oxide debris is not a significant contribution to the problem.

It is interesting to note that in one of the more recent investigations (Ref 64), it was stated that the debris had some of the characteristics of a thixotropic liquid, an observation that had been made by Godfrey in 1951 (Ref 58).

Environmental Effects

Obviously, most investigations and reported cases of fretting wear occur in a normal atmosphere where oxygen and water vapor are the main corrosive agents, although there is evidence that nitrogen may be involved in alloys containing nitride-forming elements.

Vacuum. An investigation on the effect of reduced air pressure on the fretting of a carbon steel cylinder on a flat showed that in a vacuum of 1 mPa (1.5×10^{-7} psi), the coefficient of friction was high at 3.0, and, although little measurable wear took place, considerable surface damage occurred due to adhesive transfer of metal, which ceased at pressures above 0.1 Pa (1.5×10^{-5} psi) (Ref 66). As the pressure was further increased, the coefficient of friction fell to a value of ~1.0 at 10 Pa (1.5×10^{-3} psi), whereas the wear remained low up to this pressure. Above 10 Pa (1.5×10^{-3} psi), the coefficient of friction remained constant but the wear rate increased rapidly (Fig. 25). Below this critical pressure, any debris formed was black and identified as Fe_3O_4, but above the critical pressure the debris was red and identified as α-Fe_2O_3. Similar results were obtained with an austenitic stainless steel, except in this case the plastic deformation at the low pressures resulted in martensite formation and the growth of surface fatigue cracks (Ref 67). The effects of slip amplitude and normal load were investigated on the initiation and growth of such cracks at a pressure of 4 MPa (6×10^{-7} psi); the tests showed that the number of cracks decreased rapidly at an amplitude $\geq$50 μm ($\geq$0.002 in.) but that the length of the longest crack increased markedly above this amplitude (Ref 68). The number of cracks increased steadily over the load range 8 to 30 N (0.8 to 3 kgf), whereas the length of the longest cracks was greatest at 8 N (0.8 kgf) but decreased rapidly at loads above 13 N (1.3 kgf). These results have considerable significance in relation to the fatigue aspects of fretting, particularly in connection with space research. Further work on an Fe-Cr-Ni-W amorphous alloy (hardness, 1300 HV) fretting against stainless steel gave similar results, although the coefficient of friction remained low over the whole pressure range, which was attributed to cylindrically shaped debris acting as rollers (Ref 69).

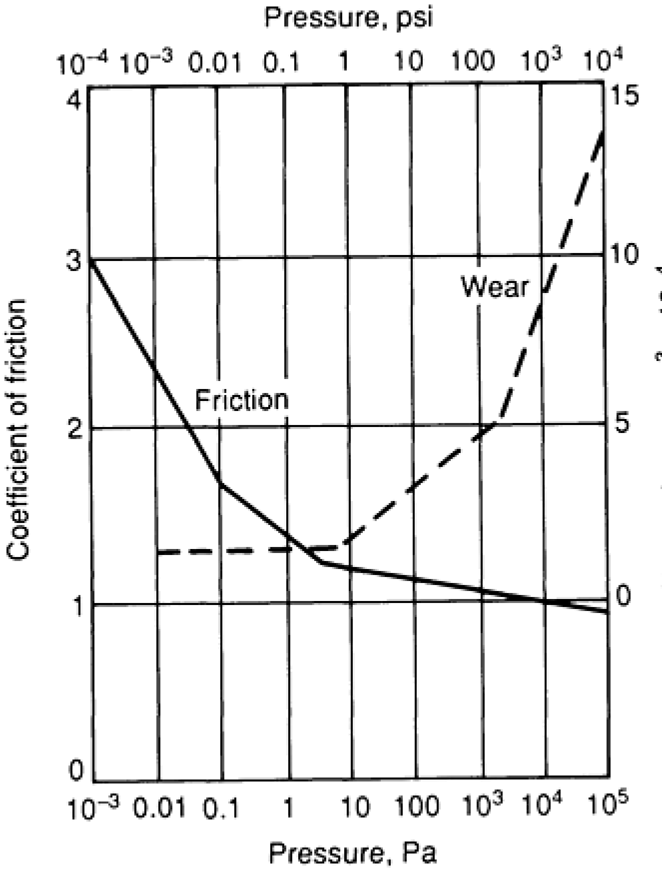


Fig. 25 Effect of atmospheric pressure on the friction and wear of a carbon steel cylinder

The effect of high pressure has not been extensively investigated. The studies that have been conducted on the effects of high pressure have been associated with the nuclear power industry, specifically relating to problems in the advanced gas-cooled reactor (AGR). In this case, the reactor is cooled by carbon dioxide at a pressure of 4 MPa (40 atm), with temperatures in the region of 600 °C (1110 °F) (see the section "Low and High Temperature" in this article).

Humidity. After oxygen, water vapor is the most significant environmental factor in fretting wear. In a study of a wide variety of pure metals tested in a hemisphere on flat configuration and at a slip amplitude of 80 μm (0.0030 in.), most metals such as silver, copper, titanium, and iron showed a peak coefficient of friction value between 10 and 15% relative humidity (RH), with a corresponding peak in wear volume (Ref 70). With chromium, the peaks appeared at 5% relative humidity, but with aluminum and nickel the peaks occurred at very low humidity. At normal atmospheric humidity (~ 50% relative humidity), all the metals except nickel showed a gradual decrease in fretting wear. This corresponds well with earlier observations on steel specimens (Ref 24). Because aluminum alloys are particularly sensitive to the effects of water vapor in fatigue situations due to the reactivity of an exposed aluminum surface and the liberation of hydrogen, further work has been devoted to the aluminum-zinc-magnesium alloy (Ref 71). Fretting wear was found to be the same in dry air as in dry argon, thus discounting oxygen as an active agent; however, in air of 60% relative humidity, the wear rate was much higher because the wear debris was nonadherent and thus metal-to-metal contact was increased. In addition, it was found that softening due to overaging of the surface material in the contact region developed in humid air.

As mentioned above, water vapor has a considerable effect on the fretting wear of certain ceramics, particularly silicon carbide and silicon nitride (Ref 56). With silicon carbide, alumina, and zirconia, the high rate of wear at 5% relative humidity is greatly reduced at humidities >50% with a steady decrease in the coefficient of friction. Silicon nitride shows a similar wear rate to silicon carbide at 5% relative humidity, but the rate increases linearly with time at ≥50% relative humidity. This is related to the propensity for silicon nitride to be transformed into silica.

Aqueous Electrolytes. Fretting of corrosion-resistant metals in aqueous environments was originally investigated because of problems associated with orthopedic implants in the human body. In some of these devices, metal-to-metal contacts (for example, screws and bone plates) could suffer fretting due to the activity of the patient. In potentiostatic

experiments on an austenitic stainless steel, it was found that the corrosion current increased linearly with the amplitude of slip at a fixed frequency and was also linearly related to frequency at a fixed amplitude (Ref 72). Where potential control was not applied, there were large drops in potential with a slow recovery when the fretting action ceased (Fig. 26). In studies in physiological saline (artificial body fluid) of fretting of type 316L stainless steel and Co-Cr-Mo alloy, the latter was found to have the lower wear rate, although the addition of albumin reduced the wear rate of the stainless steel (Ref 73). In tests with a fretting simulator using screws and a bone plate of type 316L stainless steel, the weight loss of the screws was an order of magnitude greater in 0.9% NaCl compared with the same solution to which 10% bovine serum had been added (Ref 74). Measurements of potential drop on fretting obtained in laboratory experiments were compared with *in vivo* measurements on bone plates fitted in sheep walking on a treadmill. The potential drop was found to be related to the load at low levels of torque to which the screws had been tightened (Ref 75). Further experiments on the simulator revealed that the wear weight loss of plates and screws was highest for stainless steel (2.8 mg, or 0.043 gr), less for wrought Co-Ni-Cr-Mo (0.65 mg, or 0.010 gr) and lowest for Ti-6Al-4V ELI (0.13 mg, or 0.0020 gr) (Ref 76), where ELI is extra-low interstitial. It has since been established that the effect of proteins on fretting corrosion depends on the relation of the pH of the solution to the isoelectric point of the protein. At pH values below the isoelectric point, no effect is seen, but at pH values above the isoelectric point there is a reduction in the wear rate (Ref 77).

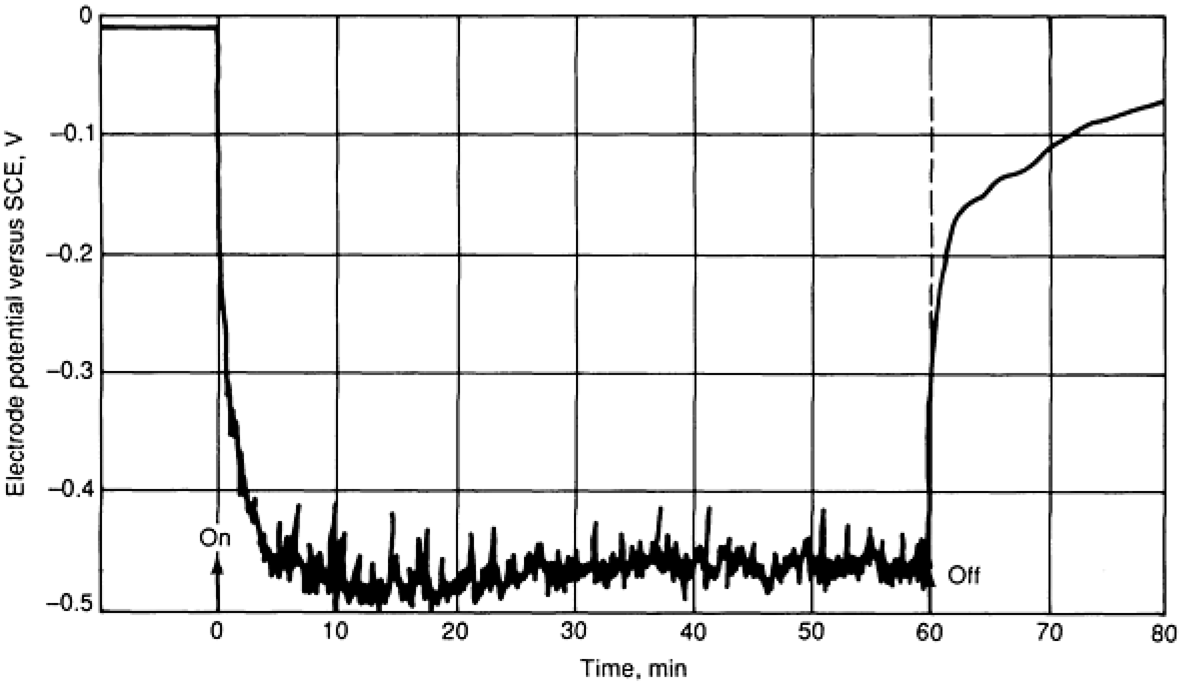


Fig. 26 Potential drop obtained when fretting Ti-6Al-4V (IMI 318) titanium alloy in a saline solution

The investigation of corrosion-resistant materials under fretting conditions in aqueous electrolytes was obviously of considerable interest because they rely on a protective oxide film for their corrosion properties. Continual disruption of this film is known to produce high corrosion rates, as confirmed by measurements of weight loss or volume of pits, which equates to rapid wear rates. The development of offshore oil rigs in the 1980s provided many more examples of fretting corrosion occurring on less corrosion-resistant materials, such as the weldable steels used in the construction of the rigs and the steel ropes used to moor them in the marine environment. The specific wear rate for a 1.5% Mn steel in air was 1.07 $m^3/N \cdot m \times 10^{-16}$ but in seawater it increased to 10.0 $m^3/N \cdot m \times 10^{-16}$ at 50 Hz and 235.5 $m^3/N \cdot m \times 10^{-16}$ at 1 Hz (Ref 78). Fortunately, application of cathodic protection of -950 mV versus standard calomel electrode (SCE) reduced the wear rate almost to that in air.

Steel ropes are extensively used in marine conditions. The development of the offshore oil industry has increased the use of these ropes as moorings for the oil rigs. The single-strand ropes are made up of layers of 5 mm (0.2 in.) diameter steel wires; successive layers are spiralled in the opposite sense to the adjacent layers. Numerous interwire contacts are referred to as trellis contacts. The fretting behavior of these wires in artificial seawater has been extensively investigated (Ref 79). In crossed-cylinder tests with an amplitude of slip of 50 μm (0.002 in.), a normal load of 3 N (0.3 kgf), and a frequency of 30 Hz, it was found that the wear volume after 5×10^6 cycles for the 0.64% C steel was 55×10^{-4} mm^3 (34×10^{-8} $in.^3$)

in air, 167 × 10^{-4} mm^3 (102 × 10^{-8} $in.^3$) in seawater, but reduced to 4 × 10^{-4} mm^3 (2 × 10^{-8} $in.^3$) when a potential of -950 mV versus SCE was applied (Ref 80). In 3.5% NaCl solution the wear was even higher at 214 × 10^{-4} mm^3 (130 × 10^{-8} $in.^3$), indicating that the magnesium salts in the seawater were having some inhibiting effect. The effect of coating the wire by hot dip galvanizing was similar to the effect of cathodic protection, reducing the wear to ~10 × 10^{-4} mm^3 (~6 × 10^{-8} $in.^3$) in seawater.

More aggressive solutions have not been widely investigated, apart from two examples by the author. Fretting of an austenitic stainless steel in 0.1 H_2SO_4 produced severe pitting in the contact region (Fig. 27). In some work related to the cooling of suspensions of radioactive solids in concentrated nitric acid, it was found that austenitic stainless steel tubes in a crossed-cylinder configuration when fretted in this mixture developed very strong adhesion as though the scouring action of the solid matter had produced clean metal surfaces that were able to form cold welds. It is interesting to note that strong adhesion was found when austenitic stainless steel was fretted against itself in liquid sodium at 500 °C (930 °F). The force to part the surfaces was calculated to require a tensile stress of 500 MPa (73 ksi), greater than the ultimate tensile strength (UTS) of the material (Ref 81).

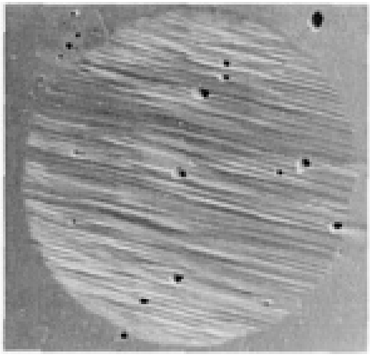


Fig. 27 Local pitting produced when an austenitic stainless steel ball is fretted against an austenitic stainless steel flat in 0.1 N H_2SO_4

Low and High Temperature. The increasing development of space engineering has meant that dynamic systems have to operate not only in vacuum but also at very low temperatures. Fretting of a spherical surface on a flat, both composed of austenitic stainless steel, at 4 K (-452 °F) in liquid helium resulted in high friction but little wear, very similar to the situation in high vacuum (see the section "Vacuum" in this article). The explanation was similar, that is, that oxide films did not grow at this temperature. At 77 K (-320 °F), the friction and wear were the same as at 293 K (-68 °F).

Much more attention has been given to high-temperature fretting in oxidizing atmospheres because this is a situation that is very common in practice (for example, in gas turbine aeroengines and in the AGR device). The pioneer work of Hurricks and Ashford (Ref 82) and later of Hurricks (Ref 83, 84) showed that there was a great reduction in the coefficient of friction and wear when mild steel was fretted in air at temperatures up to 500 °C (930 °F). In argon, however, surface damage was severe. At about this time, G.C. Wood was investigating the reciprocal sliding of certain alloys at high temperature in an oxidizing atmosphere and found that many alloys containing combinations of iron, nickel, and chromium developed a very smooth oxide, which he termed "glaze" oxide, that had low friction and low wear (Ref 85). A similar type of oxide was found in fretting a nickel-base alloy at 540 °C (1005 °F) (Ref 86). Fretting wear tests at amplitudes of 10 and 40 μm (0.0004 and 0.0016 in.) showed low coefficients of friction at 280 and 540 °C (535 and 1005 °F) compared with room temperature, and corresponding reductions in wear damage (Ref 87). However, if the air pressure is reduced to 1 mPa (1.5 × 10^{-7} psi) severe surface cracking is produced (Ref 88). The alloys that are capable of forming the "glaze" oxide are those on which an oxide with a spinel structure can develop (for example, mild steel [$FeO \cdot Fe_2O_3$], stainless steels [$FeO \cdot Cr_2O_3$, $NiO \cdot Cr_2O_3$, $NiO \cdot Fe_2O_3$], and nickel-chromium alloys [$NiO \cdot Cr_2O_3$]). If the oxide film is damaged, it is self-repairing at the high temperature but does not survive at room temperature (Ref 89). Titanium alloys do not behave in quite the same way; however, as shown in Fig. 28, implantation with bismuth or barium ions results in very low friction and wear results at temperatures between 400 and 600 °C (750 to 1110 °F) (Ref 90). For such films to survive, the substrate must have sufficient creep resistance to withstand the high stresses in the contact regions; otherwise, cracking can occur (Ref 91); see Fig. 29.

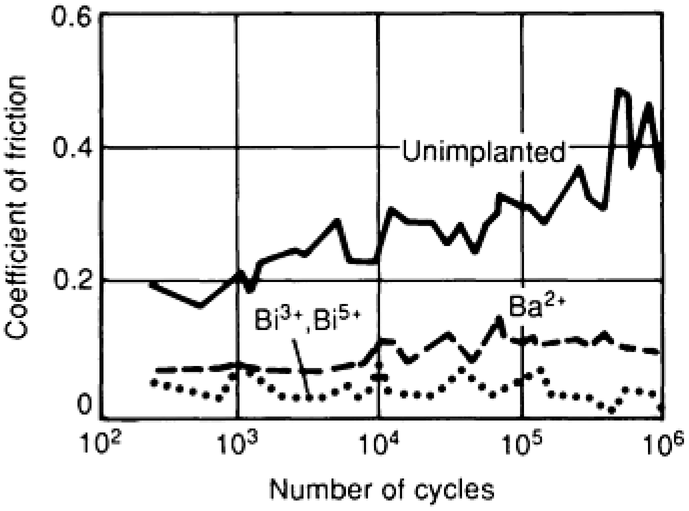


Fig. 28 Effect of ion implantation on the coefficient of friction in fretting of IMI 550 titanium alloy at 500 °C (930 °F)

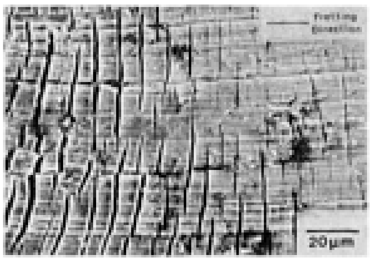


Fig. 29 Cracking of "glaze" oxide layer because of substrate creep in the fretting contact regions

Measurement of Fretting Wear

Profilometry. The very small amount of material that is removed in most fretting experiments renders the traditional method of measuring wear by weight loss impracticable. The recently developed method of surveying the surface by profilometry, where the instrument can plot out an isometric projection of the surface and the computer can produce a figure for the volume of the scar below the original surface and also the volume above the original surface, gives one of the best quantitative and descriptive assessments of the damage (Ref 92). When this method was checked with weight loss measurements, good agreement was found. Figure 30 shows the scar in a crossed-cylinder arrangement.

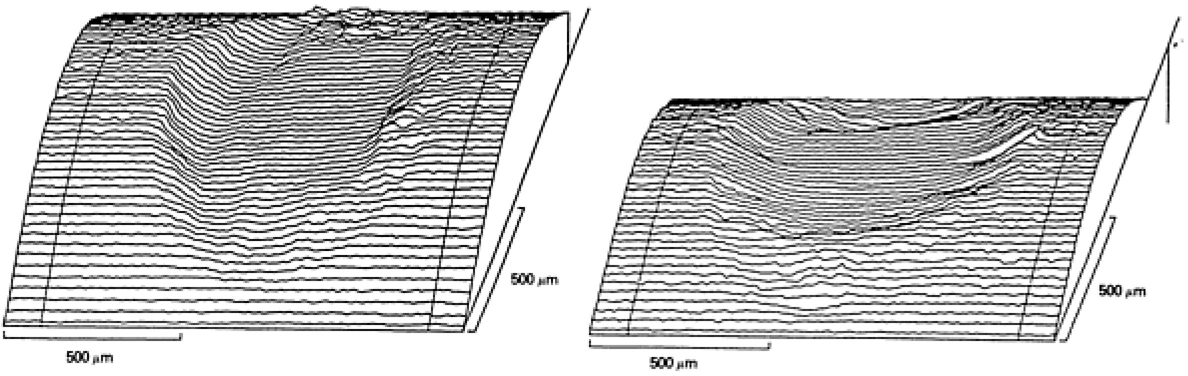


Fig. 30 Two views of an isometric projection of a scar in the crossed-cylinder arrangement

Holographic Interferometry. Another method that gives a visual impression of the damage is holographic interferometry. This enables a contour map of the surface to be generated prior to the wear process in order to compare it with a similar map generated after the wear has occurred. By using holographic image subtraction, very small changes in surface topography can be detected (Ref 93).

Thin-Layer Activation. A quantitative method applied to measuring fretting wear is "thin-layer activation" (TLA). A thin layer of radioactive atoms is produced on the surface by bombardment with a high-energy ion beam from an accelerator. After wear has taken place, debris is removed and the residual radioactivity is measured. A sensitivity of at least 0.2 μm (8 μin.) is claimed (Ref 94).

Measurement of Axial Distance. A less-sophisticated method for the case of crossed cylinders is the determination of wear from the decrease in distance between their axes (Ref 95). The advantage of this method is that it allows measurements to be made without dismantling or disturbing the contact.

Mechanism of Fretting Wear

It is generally agreed that when fretting wear is taking place between two flat surfaces or two conforming cylindrical surfaces (that is, conditions in which escape of debris is not as easy), the progress with time (or number of cycles) can be recognized as occurring in three stages (Ref 96, 97):

- Initial stage of a few thousand cycles when metal-to-metal contact is prevalent, resulting in local welding, roughening of the surface, high friction, and low contact resistance. Fatigue cracks are initiated in this stage if the movement is a result of cyclic stressing (Ref 98)
- Formation of beds of compacted oxide with a fall in coefficient of friction and erratic behavior of contact resistance as it oscillates between high and low values
- Onset of a steady state in which the friction is more or less constant and the contact resistance is generally high with occasional momentary falls to a low value

These changes are illustrated in Fig. 31, 32, and 33 and can be recognized in the wear curve (Fig. 15). The length of the initial stage depends on the amplitude of slip and the normal load (Fig. 16).

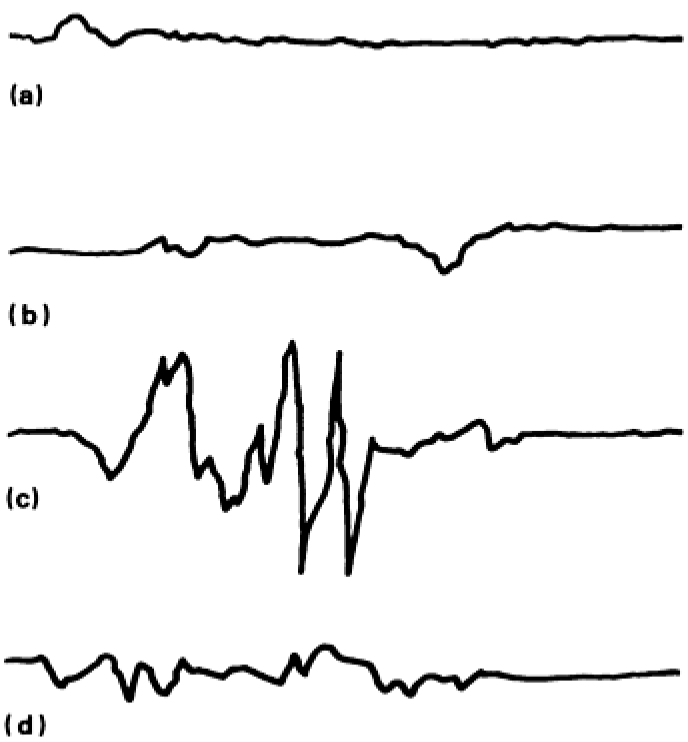


Fig. 31 Changes in surface roughness as a function of the number of fretting cycles for metal-to-metal contact. (a) As received. (b) 10^3 cycles. (c) 5×10^4 cycles. (d) 3×10^5 cycles

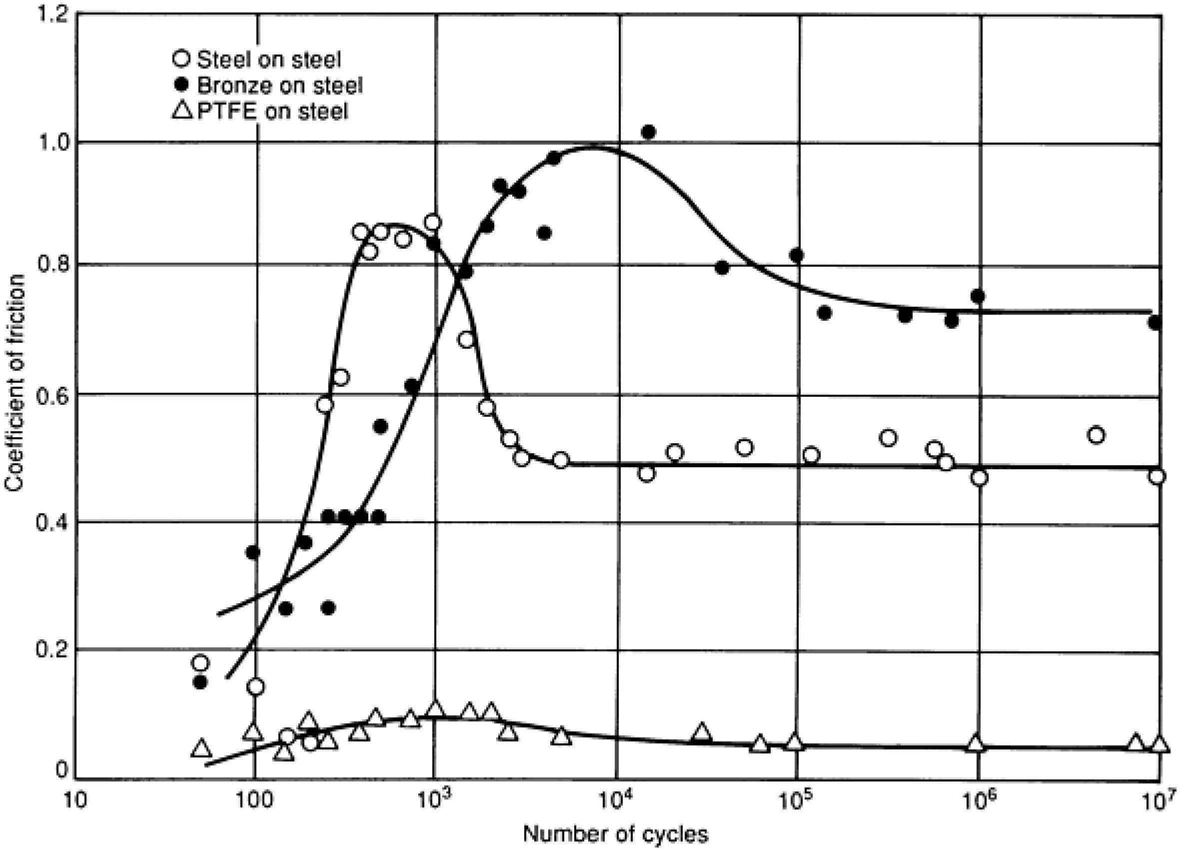


Fig. 32 Plot of the coefficient of friction versus the number of fretting cycles for three selected materials tested

on steel

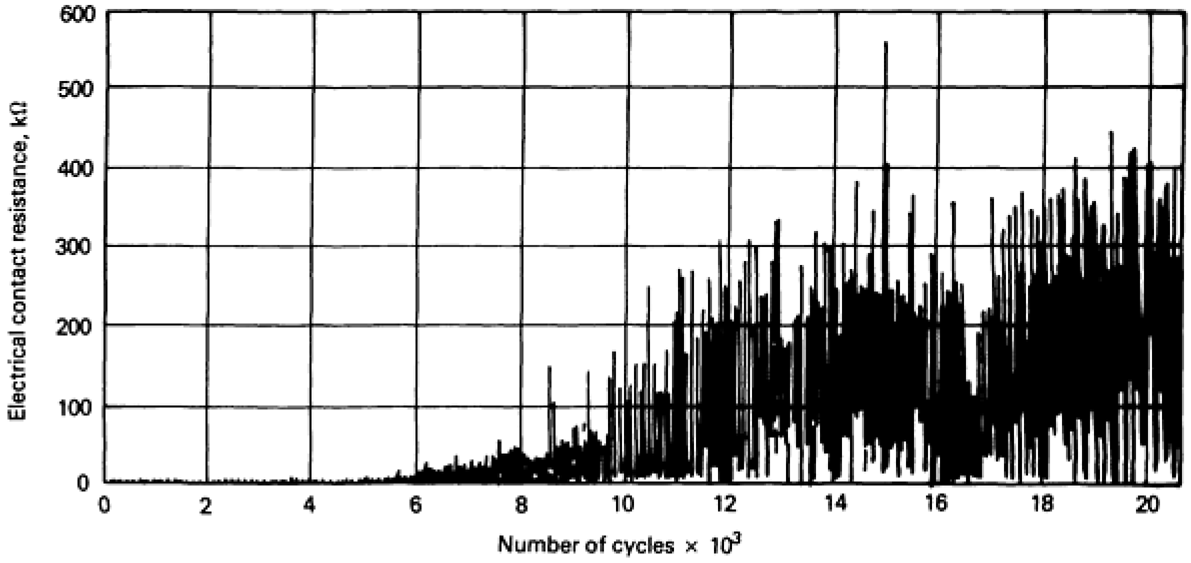


Fig. 33 Plot of contact resistance versus the number of fretting cycles

Observation of the debris on steel shows that it is platelike (Fig. 5). One of the early theories of fretting wear maintained that the debris arose by continual scraping and regrowth of oxide film (Ref 99). A modification of this idea suggests, from evidence of sections through a fretted surface on a carbon steel, that there are three well-defined zones (Ref 100):

- *Zone 1:* outer layer of compacted oxide
- *Zone 2:* severe deformation in which grains are comminuted and oriented in the fretting direction
- *Zone 3:* plastic deformation (from which it is concluded that in the steady-wear state, flake-like debris is formed by stripping oxide layers from the metal surface) (Ref 100)

The opposite point of view is that metal particles, which become progressively ground up and oxidized between the fretting surfaces, are removed. The evidence for this is that with reactive metals, such as aluminum and titanium, the debris is flakelike but contains metal (Ref 101). Furthermore, even after many millions of cycles, metal-to-metal contacts still occur in the fretting of steel flats (Ref 102). In experiments in which steel filings were introduced between two hard inert surfaces and subjected to fretting, the metal was soon converted to oxide (Ref 103). How the metal particles are generated is still a subject of controversy. However, in certain instances there is evidence that the particles are generated in the later stages of a delamination process (Fig. 34).

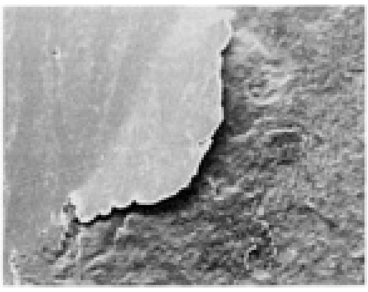


Fig. 34 Typical delamination in the fretted region produced by metal-to-metal contact

It is possible that both processes, the removal of oxide and the removal of metal, could be occurring simultaneously in different regions of the surface. In the initial stage, peaks and valleys are formed due to local welding, and these appear to persist (Ref 104). The in-between areas become filled with compacted debris that has a layerlike structure, from which oxide flakes can become detached. These areas of compacted debris are maintained by metal particles produced by the continuing metal-to-metal contacts.

Although the oxide debris is usually harder than the metal from which it arises, abrasion by the oxide is not a contributory factor on steels to the wear. In fact, the debris, by limiting metal-to-metal contact and possibly taking up some of the movement between the surfaces, has more of a protective effect (Ref 60).

Prevention of Fretting Damage

The steps that can be taken to reduce or to eliminate the damage due to fretting are extremely diverse, and each case or prospective case needs to be analyzed individually in order to select the most promising method to be applied. In this article, it is only possible to give some general indications of how the problem can be tackled. Beard (Ref 105) has pointed out that it is useful to know whether the movement is "force controlled" or whether it is "displacement controlled." Force-controlled fretting is experienced in the case of partial slip; therefore, increasing the coefficient of friction or increasing the normal load will reduce the slip. In displacement-controlled fretting, where slip is occurring over the whole interface, the amplitude is constant and wear will be reduced by reducing the normal load and also the coefficient of friction.

Improved Design. Because fretting only arises where there is relative movement between two surfaces, the elimination of this movement is a prime objective. Fretting often arises as a result of a stress concentration. In any case, reducing the stress via improved design (for example, raised wheel seats or stress-relieving grooves in hub/axle assemblies) is demanded by fatigue considerations. Increasing the normal pressure by reducing the area of contact (when the normal load is to be kept constant) or increasing the normal load will reduce the area of slip, but such action may introduce fatigue problems. The ultimate goal is to make the component in one piece and to get rid of the junction, but this is not always possible.

Surface Finish. Rough surfaces are less prone to fretting damage than highly polished surfaces. Rough surfaces can be obtained by shot peening with glass beads or steel shot. Shot peening has the advantage of work hardening the surface, although the residual compressive stress appears to have little effect on the wear process; however, it has a considerable effect if fatigue crack initiation and propagation is involved. Shot peening is often applied as a surface preparation for the application of coatings.

Coatings. Surface treatments that radically change the chemical composition of the surface can be divided into three categories:

- Methods where foreign atoms or ions are introduced into the existing surface by diffusion (for example, carburizing, nitriding, chromizing, sheradizing, and aluminizing) or by bombardment (for example, ion implantation)
- Methods where the surface reacts with a chosen environment to form a compound (for example, oxidation, anodizing, and phosphating)
- Methods where an entirely foreign material is applied to the surface (for example, electrodeposition, plasma spray, ion plating, physical vapor deposition [PVD], and chemical vapor deposition [CVD])

Most of these coatings are harder and consequently more brittle than the substrate, which, if it is not strong enough, will plastically deform in the fretting contact, where local stresses can be high. This plastic deformation will crack and break up the coating. If the coating is particularly hard (for example, titanium nitride, with a hardness of 300 HV), further damage can occur by abrasion (Ref 106).

Carburizing and nitriding of steels is a well-established treatment for reducing wear in gears and is also effective against fretting (Ref 107). Electrodeposited coatings are extensively used in electrical contacts; gold alloys are the preferred coating because they are not subject to oxide formation, which is disastrous in its effect on conductivity (Ref 108).

The growth of oxide films at high temperatures has been specified above to provide wear-resistant self-repairing coatings. These have been improved by ion implantation, particularly in titanium alloys, to give coatings that have low friction and low wear (Ref 90).

Anodizing is applied to aluminum alloys and the hard coating provides protection against fretting, but it may cause wear of the opposing surface (Ref 109).

Inserts. Separation of the two surfaces by an insert, either a shim of a soft metal or a shim of a polymer with a low elastic modulus, will sometimes be effective. The intention is to take up the movement by either plastic deformation or elastic deformation of the material. A combination of metal and polymer can be more effective because it combines the advantages of both materials--the low elasticity of the polymer and the good conductivity of the metal. Bronze-filled polytetrafluorethylene (PTFE) when fretted against steel has a very low friction coefficient and undetectable wear (Ref 110).

Lubricants. Lubricants are usually difficult to apply in fretting contacts because of the difficulty of maintaining the lubricant in the contact, but lubricants are included in the construction of steel ropes (for example, locked coil ropes) (Ref 111). In a survey of the effect of additives to a base mineral on the fretting wear of steel surfaces, zinc di-*n*-octyldithiophosphate was found to be most effective (Ref 112). Solid lubricants applied to the surfaces (for example, MoS_2) reduce the coefficient of friction initially, but eventually tend to wear away (Ref 113).

References

1. E.M. Eden, W.N. Rose, and F.L. Cunningham, Endurance of Metals, *Proc. Inst. Mech. Eng.,* Vol 4, 1911, p 839-974
2. G.A. Tomlinson, The Rusting of Steel Surfaces in Contact, *Proc. R. Soc. (London) A,* Vol A115, 1927, p 472-483
3. E.J. Warlow-Davies, Fretting Corrosion and Fatigue Strength, *Proc. Inst. Mech. Eng.,* Vol 146, 1941, p 32-38
4. J.R. McDowell, Fretting Corrosion Tendencies of Several Combinations of Materials, *Symposium on Fretting Corrosion,* STP 144, American Society for Testing and Materials, 1953, p 2439
5. A.J. Fenner and J.E. Field, La Fatigue dans les Conditions de Frottement, *Rev. Métall.,* Vol 55, 1958, p 475-485
6. K.T. Styles, The Influence of Fret Corrosion upon the Fatigue Life of Machine Components, *Met. Australas.,* Mar 1980, p 4-6
7. J.F. Warriner, "Thin Shell Bearings for Medium Speed Diesel Engines," No. 364, Diesel Engineers and Users Association, Feb 1975
8. R. Heinz, Fretting Wear--Examples, Test Methods and Preventive Measures, *Mater. Wiss. Werkstofftech.,* Vol 20, 1989, p 14-20
9. R.G. Mitchell, An Airline View of the Corrosion Problem, *Ind. Corr.,* July 1983, p 11-17
10. P. Greenfield and R.W. Suhr, The Factors Affecting the High-Cycle Fatigue Strength of Low-Pressure Turbine and Generator Rotors, *Proceedings of International Conference on Fatigue of Engineering Materials and Structures,* 15-19 Sept 1986 (Sheffield), *Inst. Mech. Eng.,* Vol 1, 1986, p 165-186.
11. H.W. Fricker, Fretting in Tube Supports of Heat Exchangers, *Proceedings of Conference on Component Design in High Temperature Reactors Using Helium as a Coolant,* Nuclear Engineering Group of Institution of Mechanical Engineers, 3-4 May 1972, p 213-231
12. H.J. Connors, Flow-Induced Vibration and Wear of Steam Generator Tubes, *Nucl. Technol.,* Vol 55, Nov 1981, p 311-331
13. P.J.W. Bongers, P.L.M. Dobs, and T.J.M. Linssen, Fatigue Crack Initiation on the Outside Surface of High-Pressure Tubes Caused by Tube Support, Proceedings of Conference on Safety in High Pressure Polyethylene Process, Symposium III, Mar 1977 (Houston), American Institute of Chemical Engineers
14. M. Antler, Electrical Effects of Fretting Connector Contact Materials: A Review, *Wear,* Vol 106, 1985, p 5-33

15. P.J. Kennedy, M.B. Peterson, and L. Stallings, An Evaluation of Fretting at Small Slip Amplitudes, *Materials Evaluation under Fretting Conditions,* STP 780, American Society for Testing and Materials, 1982, p 30-48
16. P.J. Kennedy, L. Stallings, and M.B. Peterson, A Study of Surface Damage at Low-Amplitude Slip, *ASLE Trans.,* Vol 27 (No. 4), 1984, p 305-312
17. K.L. Johnson, Surface Interaction between Elastically Loaded Bodies under Tangential Forces, *Proc. R. Soc. (London) A,* Vol 230, 1955, p 531-548
18. D. Nowell and D.A. Hills, Crack Initiation Criteria in Fretting Fatigue, *Wear,* Vol 136, 1990, p 329-343
19. M. Kuno, "Fretting Damage of High Carbon Chromium Steel," Ph.D. thesis, University of Nottingham, 1988
20. S. Goto and R.B. Waterhouse, Fretting and Fretting-Fatigue of Titanium Alloys under Conditions of High Normal Load, *Proceedings of the 4th International Conference on Titanium,* TMS/AIME, Vol 3, 1980, p 1337-1384
21. O. Vingsbo and S. Söderberg, On Fretting Maps, *Wear,* Vol 126, 1988, p 131-147
22. U. Bryggman and S. Söderberg, Contact Conditions in Fretting, *Wear,* Vol 110, 1986, p 1-17
23. F.P. Bowden and D. Tabor, "The Friction and Lubrication of Solids," Oxford University Press, 1950, p 19
24. I.M. Feng and H.H. Uhlig, Fretting Corrosion of Mild Steel in Air and Nitrogen, *J. Appl. Mech.,* Vol 21, 1954, p 395-400
25. R.B. Waterhouse, Fretting, *Treatise on Materials Science and Technology,* D. Scott, Ed., Academic Press, 1978, p 259-286
26. S. Söderberg, U. Bryggman, and T. McCullough, Frequency Effects in Fretting Wear, *Wear,* Vol 110, 1986, p 19-34
27. H. Kongsjorden, J. Kulsetas, and J. Sletbak, "Degradation of Electrical Contacts Caused by Oscillatory Micromotion between the Contact Members," Proceedings of the 9th International Conference on Electrical Contact Phenomena, 11-15 Sept 1978 (Chicago)
28. M. Antler, Electrical Effects of Fretting Connector Contact Materials: A Review, *Wear,* Vol 106, 1985, p 5-33
29. M. Braunovic, M.S. McIntyre, W.J. Chaurion, and I. Aitchison, Surface Analysis of Fretting Damage in Electrical Contacts of Aluminum with Different Contacting Materials, *Electrical Contacts 1983,* Proceedings of the 29th Holm Conference on Electrical Contacts (Chicago), Illinois Institute of Technology, p 232-242
30. J.J. Mottine and B.T. Reagor, Investigation of Fretting Corrosion at Dissimilar Metal Interfaces in Socketed IC Device Applications, *Electrical Contacts 1983,* Proceedings of the 29th Holm Conference on Electrical Contacts (Chicago), Illinois Institute of Technology, p 67-71
31. M. Kuno and R.B. Waterhouse, The Effect of Oscillatory Direction on Fretting Wear under Crossed Cylinder Contact Conditions, *Eurotrib 89,* Proceedings of the 5th International Congress on Tribology, 12-15 June 1989 (Helsinki), Vol 3, p 30-35
32. K.B. Elliott, H.H. Mabie, M.J. Furey, and L.D. Mitchell, A Vibration Analysis of a Bearing/Cartridge Interface for a Fretting Corrosion Study, *J. Lubr. Technol. (Trans. ASME),* Vol 105, 1983, p 518-525
33. N.L. Golego, M.N. Rozhkov, V.P. Onoprienko, A.S. Nakrakui, and N.V. Kozii, Wear of Materials in Fretting Corrosion in Connection with Conditions of Vibrocontact Interaction, *Fiz. Khim. Mekh. Mater.,* Vol 21, 1985, p 71-75
34. J.H. Cha, M.W. Wambsganss, and J.A. Jendrzejcyk, Experimental Study on Impact/Fretting Wear in Heat Exchanger Tubes, *J. Pressure Vessel Technol. (Trans. ASME),* Vol 109, 1987, p 265-274
35. A.W.J. de Gee, C.P.L. Commissaris, and J.H. Zaat, The Wear of Sintered Aluminum Powder (SAP) under Directions of Vibrational Contact, *Wear,* Vol 7, 1964, p 535-550
36. P.L. Ko, Experimental Studies of Tube Fretting in Steam Generators and Heat Exchangers, *J. Pressure Vessel Technol. (Trans. ASME),* Vol 101, 1979, p 125-133
37. H.J. Connors, Flow-Induced Vibration and Wear in Steam Generator Tubes, *Nucl. Technol.,* Vol 55, 1981,

p 311-331

38. S.F. Calhoun, "Effects of Metal Hardness and Surface Finish upon Fretting Corrosion," Technical Report No. 63-1835, Rock Island Arsenal, 28 May 1963
39. R.B. Waterhouse, Residual Stresses and Fretting, *International Guidebook on Residual Stresses,* A. Niku-Lari, Ed., Pergamon, Oxford, 1987, p 511-525
40. C.A. Bergman, R.C. Cobb, and R.B. Waterhouse, The Effect of Shot-Peening on the Friction and Wear of a Carbon Steel in Fretting Conditions, *Wear of Materials,* American Society of Mechanical Engineers, 1987, p 33-37
41. J. Labedz, Effect of Residual Stresses on Wear of the Surface Layer of Components Operating under Fretting Conditions, *Metal Treatments against Wear, Corrosion, Fretting and Fatigue,* A. Niku-Lari and R.B. Waterhouse, Ed., Pergamon, Oxford, 1988, p 87-98
42. K. Endo, Practical Observations of Initiation and Propagation of Fretting Fatigue Cracks, *Fretting Fatigue,* R.B. Waterhouse, Ed., Applied Science Publishers, London, 1981, p 127-141
43. A. Yournadj, G. Maeder, and A. Moison, Geometric and Mechanical Aspects of Fretting Corrosion, *Proceedings of the 4th European Tribology Congress,* 9-12 Sept 1985 (Euiby, France), Vol 2, p 1-9
44. G. Fair, B. Noble, and R.B. Waterhouse, The Stability of Compressive Stresses Induced by Shot-Peening under Conditions of Fatigue and Fretting-Fatigue, *Advances in Surface Treatments,* Vol I, A. Niku-Lari, Ed., Pergamon, Oxford, 1984, p 3-8
45. M. Kuno, R.B. Waterhouse, D. Nowell, and D.A. Hills, Initiation and Growth of Fretting Fatigue Cracks in the Partial Slip Regime, *Fatigue Fract. Eng. Mater. Struct.,* Vol 12, 1989, p 387-398
46. T. Kayaba and A. Iwabuchi, Effect of the Hardness of Hardened Steels and the Action of Oxides on Fretting Wear, *Wear,* Vol 66, 1981, p 27-41
47. O. Vingsbo, M. Odfalk, and Ning-E Shen, Fretting Maps and Fretting Behaviour of Some FCC Metal Alloys, *Wear,* Vol 13, 1990, p 153-167
48. J. Dobromirski and I.O. Smith, Metallographic Aspects of Surface Damage, Surface Temperature and Crack Initiation in Fretting Fatigue, *Wear,* Vol 117, 1987, p 347-357
49. J. Beard and P.F. Thomason, The Mechanics of Fatigue Crack Nucleation in a Low Alloy Steel under Fretting Conditions, *Fatigue 87,* Conference Proceedings, 28 June to 3 July 1987 (Charlottesville, VA), Vol 1, 1987, p 53-62
50. R. Becker and G. Sepold, Fretting of Amorphous Alloys, *Rapidly Solidified Amorphous and Crystalline Alloys,* B.H. Kear, B.C. Giessen, and M. Cohen, Ed., Elsevier Science Publishing, 1982, p 295-299
51. J. Sato, M. Sato, and S. Yamamoto, Fretting Wear of Stainless Steel, *Wear,* Vol 69, 1981, p 167-177
52. P.A. Higham, F.H. Stott, and B. Bethune, Mechanisms of Wear of the Metal Surface during Fretting Corrosion of Steel on Polymers, *Corros. Sci.,* Vol 18, 1978, p 3-13
53. P.A. Higham, F.H. Stott, and B. Bethune, The Influence of Polymer Composition on the Wear of the Metal Surface during the Fretting of Steel on Polymer, *Wear,* Vol 47, 1978, p 71-80
54. O. Jacobs, K. Friedrich, G. Marom, K. Schulte, and H.D. Wagner, Fretting Wear Performance of Glass-Carbon and Aramid Fibre/Epoxy- and PEEK-Composites, *Wear of Materials,* 1989, American Society of Mechanical Engineers, 1989, p 495-500
55. M. Kuno, R.B. Waterhouse, and B.R. Pearson, Fretting Wear in Sintered Alumina and Tungsten Carbide Cermet-Metal Couples, *Wear of Materials,* 1987, American Society of Mechanical Engineers, 1987, p 371-380
56. D. Klaffke, Fretting Wear of Ceramics, *Tribol. Int.,* Vol 22, 1989, p 89-101
57. D.R. Horn, R.B. Waterhouse, and B.R. Pearson, Fretting Wear of Sintered Alumina, *Wear,* Vol 113, 1986, p 225-232
58. D. Godfrey, "Investigation of Fretting by Microscopic Observation," Report No. 1009, NACA, 1951
59. J. Sato, M. Shima, and T. Sugiwara, A Fundamental Study of Fretting Damage to Glass Using an Improved Apparatus, *Wear,* Vol 106, 1985, p 53-61
60. M. Godet, The Third Body Approach: Mechanical View of Wear, *Wear,* Vol 100, 1984, p 437-452

61. J.F. Andrews, P.D. Donovan, and J. Stringer, Fretting Corrosion Products of Aluminium Alloys, *Br. Corros. J.,* Vol 3, 1968, p 85-87
62. C. Colombié, Y. Berthier, A. Floquet, L. Vincent, and M. Godet, Fretting: Load Carrying Capacity of Wear Debris, *J. Tribol. (Trans. ASME),* Vol 106, 1984, p 194-201
63. D. Aldham, J. Warburton, and R.E. Pendlebury, The Unlubricated Fretting Wear of Mild Steel in Air, *Wear,* Vol 106, 1985, p 177-201
64. J. Warburton, The Fretting of Mild Steel in Air, *Wear,* Vol 131, 1989, p 365-386
65. R.E. Pendlebury, Unlubricated Fretting of Mild Steel in Air at Room Temperature, Pt. II: Electrical Contact Resistance Measurements and the Effect of Wear on Intermittent Loading, *Wear,* Vol 118, 1987, p 341-364
66. A. Iwabuchi, T. Kayaba, and K. Kata, Effect of Atmospheric Pressure on Friction and Wear of 0.45%C Steel in Fretting, *Wear,* Vol 91, 1983, p 289-305
67. A. Iwabuchi, K. Kato, and T. Kayaba, Fretting Properties of SUS304 Stainless Steel in Vacuum Environment, *Proceedings of the JSLE Tribology Conference,* 8-10 July 1985 (Tokyo), Vol 1, Japan Society of Lubrication Engineers, 1985, p 29-34
68. K. Kato, T. Suzuki, A. Iwabuchi, and K. Hokkirigawa, Fretting Damage of SUS304 in a Vacuum, *JSME Int. J.,* Vol 30, 1987, p 330-336
69. A. Iwabuchi, H. Matsuzaki, and K. Hori, The Effect of Ambient Pressure on the Tribological Properties of an Amorphous Alloy in Fretting, *Wear of Materials 1989,* American Society of Mechanical Engineers, 1989, p 595-600
70. H. Goto and D.H. Buckley, The Influence of Water Vapour in Air on the Friction Behaviour of Pure Metals during Fretting, *Tribol. Int.,* Vol 18, 1985, p 237-245
71. H. Goto, M. Ashida, and K. Endo, The Influence of Oxygen and Water Vapour on the Friction and Wear of an Aluminum Alloy under Fretting Conditions, *Wear,* Vol 116, 1987, p 141-155
72. M.P. Sherwin, D.E. Taylor, and R.B. Waterhouse, An Electrochemical Investigation of Fretting Corrosion in Stainless Steel, *Corros. Sci.,* Vol 11, 1971, p 419-429
73. S.D. Cook, G.J. Gianoli, A.J.T. Clemow, and R.J. Haddad, Fretting Corrosion in Orthopaedic Alloys, *Biomat. Med. Dev. Art. Org.,* Vol 11, 1983-84, p 281-292
74. S.A. Brown and K. Merritt, Fretting Corrosion in Saline and Serum, *J. Biomed. Mater. Res.,* Vol 15, 1981, p 479-488
75. S.A. Brown and J.P. Simpson, Crevice and Fretting Corrosion of Stainless Steel Plates and Screws, *J. Biomed. Mater. Res.,* Vol 15, 1981, p 867-878
76. S.A. Brown and K. Merritt, Fretting Corrosion of Plates and Screws: An in vitro Test Method, *Corrosion and Degradation of Implant Materials: Second Symposium,* STP 859, A.C. Fraker and C.D. Griffin, Ed., ASTM, 1985, p 105-116
77. K. Merritt and S.A. Brown, Effect of Proteins and pH on Fretting Corrosion and Metal Ion Release, *J. Biomed. Mater. Res.,* Vol 22, 1988, p 111-120
78. M.P. Overs and R.B. Waterhouse, Fretting Wear of Two Weldable Structural Steels (HY80 and 50D) in Seawater, *Wear of Materials 1983,* American Society of Mechanical Engineers, 1983, p 541-545
79. B.R. Pearson and R.B. Waterhouse, The Fretting Corrosion in Seawater of Materials Used in Offshore Structures, *Proceedings of the 9th International Congress on Metallic Corrosion,* 5-9 June 1984 (Toronto), Vol 2, NRC Canada, 1984, p 334-341
80. B.R. Pearson, P.A. Brook, and R.B. Waterhouse, Fretting in Aqueous Media, Particularly of Roping Steels in Seawater, *Wear,* Vol 106, 1985, p 225-260
81. M.W.J. Lewis and C.S. Campbell, Fretting and Wear of Stainless and Ferritic Steels in Liquid-Metal Fast Breeder Reactor Steam Generators, *Nucl. Technol.,* Vol 55, 1981, p 460-469
82. P.L. Hurricks and K.S. Ashford, The Effect of Temperature on the Fretting Wear of Mild Steel, *Proc. Inst. Mech. Eng.,* Vol 184, Pt. 3L, 1969-70, p 165-175
83. P.L. Hurricks, The Fretting Wear of Mild Steel from Room Temperature to 200 °C, *Wear,* Vol 19, 1972, p

207-229

84. P.L. Hurricks, The Fretting Wear of Mild Steel from 200 to 500 °C, *Wear,* Vol 30, 1974, p 189-212
85. F.H. Stott, D.S. Lui, and G.C. Wood, The Structure and Mechanism of the "Glaze" Oxide Layers Produced on Nickel-Based Alloys during Wear at High Temperatures, *Corros. Sci.,* Vol 13, 1973, p 449-469
86. M.M. Hamdy and R.B. Waterhouse, The Fretting-Fatigue Behaviour of a Nickel-Based Alloy (Inconel 718) at Elevated Temperatures, *Wear of Materials 1979,* American Society of Mechanical Engineers, 1979, p 351-355
87. M.M. Hamdy and R.B. Waterhouse, The Fretting Wear of Ti-6Al-4V and Aged Inconel 718 at Elevated Temperatures, *Wear,* Vol 71, 1981, p 237-248
88. A. Iwabuchi, Fretting Wear of Inconel 625 at High Temperature and in High Vacuum, *Wear,* Vol 106, 1985, p 163-175
89. M.M. Hamdy and R.B. Waterhouse, The Stability of the Protective Glaze Oxide Formed during High Temperature Fretting of an Aged Nickel-Based Alloy, Inconel 718, *Wear of Materials 1983,* American Society of Mechanical Engineers, 1983, p 546-549
90. A. Iwabuchi and R.B. Waterhouse, The Effect of Ion Implantation on the Fretting Wear of Four Titanium Alloys at Temperatures up to 600 °C, *Wear of Materials 1985,* American Society of Mechanical Engineers, 1985, p 471-476
91. M.M. Hamdy, B. Noble, and R.B. Waterhouse, Fatigue Crack Initiation in IMI829 Caused by High Temperature Fretting, *J. Mater. Sci.,* Vol 18, 1983, p 493-502
92. D.G. Chetwynd, F.A. McKee, and C.H. Wride, The Measurement of Wear Damage, *Proc. Inst. Mech. Eng.,* Vol 201 4C, 1987, p 251-258
93. J.T. Atkinson and M.J. Lalor, Holographic Interferometry Applied to Minimal Wear Measurement, 1st European Congress, Appl. to Metrology, *SPIE J.,* Vol 136, 1977, p 107-113
94. T.W. Conlon, The Scope for TLA in Corrosion and Wear Studies, *Metall. and Mater. Technol.,* Vol 2, 1979, p 143-146
95. J. Warburton and R. Bradford, The Progressive Wear of Tubes: The Volumes of the Intersections of Cylinders with Each Other and with Flats, *Wear,* Vol 110, 1986, p 331-352
96. P.L. Hurricks, The Mechanism of Fretting--A Review, *Wear,* Vol 15, 1970, p 389-409
97. D. Aldham, J. Warburton, and R.E. Pendlebury, The Unlubricated Fretting of Mild Steel in Air, *Wear,* Vol 106, 1985, p 177-201
98. R.B. Waterhouse, The Role of Adhesion and Delamination in the Fretting Wear of Metallic Materials, *Wear,* Vol 45, 1977, p 355-364
99. H.H. Uhlig, Mechanism of Fretting Corrosion, *J. Appl. Mech.,* Vol 21, 1954, p 401-407
100. X. Zhang, C. Zhang, and C. Zhu, Slip Amplitude Effects and Microstructural Characteristics of Surface Layers in Fretting Wear of Carbon Steel, *Wear,* Vol 134, 1989, p 297-309
101. R.B. Waterhouse, Fretting Corrosion and Fretting Fatigue in the Transport Industry, Proceedings of the 6th European Congress on Metallic Corrosion, 19-23 Sept 1977 (London)
102. R.E. Pendlebury, Unlubricated Fretting and Sliding Wear of Steels in Air, *Tribology--Friction, Lubrication and Wear, Fifty Years On,* Conference Proceedings, 1-3 July 1987 (London), Institute of Mechanical Engineers, Vol 1, 1987, p 267-275
103. R.E. Pendlebury and P.A. Tempest, "Unlubricated Fretting Wear of Mild Steel Surfaces in Air at Room Temperature, Part III: The Conversion of (Artificial) Metallic Wear Particles to Oxide," Report TPRD/B/0985/R87, AGR/CIF/TWG/P(87)232, Central Electricity Generating Board, Oct 1987
104. R.E. Pendlebury, "Unlubricated Fretting Wear of Mild Steel Surfaces in Air at Room Temperature," Report TPRD/B/0674/N85, AGR/CIF/TWG/P(85)164, Central Electricity Generating Board, June 1985
105. J. Beard, The Rational Selection of Palliatives for Avoidance of Fretting, *Tribology--Friction, Lubrication and Wear, Fifty Years On,* Conference Proceedings, 1-3 July 1987 (London), Institute of Mechanical Engineers, Vol 1, 1987, p 311-319
106. R.C. Cobb and R.B. Waterhouse, The Fretting Wear of Certain Hard Coatings Including TiN Applied to a

0.4C Steel, *Tribology--Friction, Lubrication and Wear, Fifty Years On,* Conference Proceedings, 1-3 July 1987 (London), Institute of Mechanical Engineers, Vol 1, 1987, p 303-310

107. J.C. Gregory, A Salt Bath Treatment to Improve the Resistance of Ferrous Metals to Scuffing, Wear, Fretting, and Fatigue, *Wear,* Vol 9, 1966, p 249-281
108. J.W. Souter and W. Staunton, The Fretting Wear of Electrodeposited Contact Coatings, *Trans. Inst. Met. Finish.,* Vol 66, 1988, p 8-14
109. R.B. Waterhouse, The Effect of Surface Treatments on the Fretting Wear of an Aluminum Alloy (RR58)/Steel (BS970 080 M40) Couple, *Proceedings of Conference on Surface Engineering, Current Trends and Future Prospect,* 25-27 June 1990 (Toronto), S.A. Meguid, Ed., Elsevier Applied Science, 1990, p 325-334
110. D.J. Varley and R.B. Waterhouse, Fretting of Bronze-Filled PTFE against Mild Steel, *Proceedings of Japanese International Tribology Conference* (Nagoya), 29 Oct to 1 Nov 1990, Japanese Society of Tribologists, Vol 2, 1990, p 821-825
111. D.E. Taylor and R.B. Waterhouse, Fretting Fatigue in Steel Ropes, *Lubr. Eng.,* Vol 27, 1971, p 123-127
112. J. Sato, M. Shima, T. Sugiwara, and A. Tahara, Effect of Lubricants on Fretting Wear of Steel, *Wear,* Vol 125, 1988, p 83-95
113. D. Godfrey and E.E. Bisson, "Bonding of Molybdenum Disulphide to Various Materials to Form a Solid Lubricating Film, Pt. II. Friction and Endurance Characteristics of Films Bonded by Practical Methods," Technical Note No. 2802, NACA, 1952

Rolling Contact Wear

Peter J. Blau, Metals and Ceramics Division, Oak Ridge National Laboratory

Introduction

ROLLING CONTACT WEAR (RCW) results from the repeated mechanical stressing of the surface of a body rolling on another body. Only a brief overview of the subject is presented here. To obtain more comprehensive and detailed information, especially about failure analysis and materials selection for rolling contact wear, the reader is referred to the references cited in the relevant subsections. The wear can occur either on the roller surface or on the rolled surface. Rolling contact wear is a major cause of failure in rolling-element bearings, gear teeth, wheel-rail contacts, roller guides in production machinery, cam roller followers in internal combustion engines, and rolling mill rolls. It is a particularly insidious form of wear, because it is sometimes difficult to detect in its early stages. The precursor flaws may be hidden from view as they grow below the contact surface. By the time cracks grow large enough to emerge at the surface and produce wear particles or delaminations, these particles may become large spalls or flakes, resulting in immediate component loss of function or efficiency.

The first signs of RCW may include any of the following:

- Progressive noise in a bearing (rattling or grinding)
- Progressive heating in a bearing
- Onset of vibration in a precision bearing
- Rapid rise in wear debris content in oil samples (later stages of deterioration)
- Appearance of periodic surface irregularities in rolled sheet products

After a warning signal occurs, disassembly of suspect bearings, gears, and rolling contact components will then reveal the nature and extent of damage. Because RCW is generally produced by repetitive mechanical stressing of the affected surface(s), it is often associated with, or even referred to, as rolling contact fatigue (RCF). For the purposes of this

discussion, the difference between RCW and RCF is that, in RCF, surface fatigue is the damage accumulation process that eventually results in wear particle formation. Rolling contact fatigue may continue for hundreds, thousands, or even millions of cycles before the first wear particles are removed. Furthermore, the corners of pits or other RCW damage features may act as nucleation sites for additional fatigue cracks and spread the damage across the surface. Because RCF and RCW are so closely related, the causes and effects of both processes will be discussed.

The magnitude of the effects of RCW varies from one tribosystem to another. Sometimes, a component can sustain appreciable RCW damage before its function is impaired; other times, loss of performance immediately results from the first spall. For example, a guide roller in a hot metal bar-handling system may sustain considerable RCW, but this wear may be unimportant as long as the component continues to function adequately. In the case of ultraprecision ball bearings for missile guidance systems, however, a very small spall may cause the center of mass of the rapidly spinning bearing to shift, resulting in significant guidance errors.

Rolling contact is frequently accompanied by slip or sliding. The complex motions experienced by tribocomponents in many types of rolling contact situations produce at least a small percentage of slip or sliding. Pure rolling is probably the exception rather than the rule in the diverse applications of rolling components. For example, Fig. 1 is a schematic profile of meshing spur gears. When the gear teeth first touch, there is a measure of sliding plus rolling. When the contact point coincides with the pitch circle, there is pure or nearly pure rolling (depending on the accuracy of gear alignment or lateral vibration). Past this point, slip again occurs between tooth surfaces; it reaches its second maximum just at the point where the surfaces separate. Slip can result in scuffing or adhesive wear damage to the mating surfaces if the lubrication is inadequate. Therefore, proper lubricant and surface treatment selection is important to minimize the deleterious effects of slip in many types of rolling contact arrangements. Lubrication and wear of rolling-element bearings is discussed in the article "Friction and Wear of Rolling-Element Bearings" in this Volume.

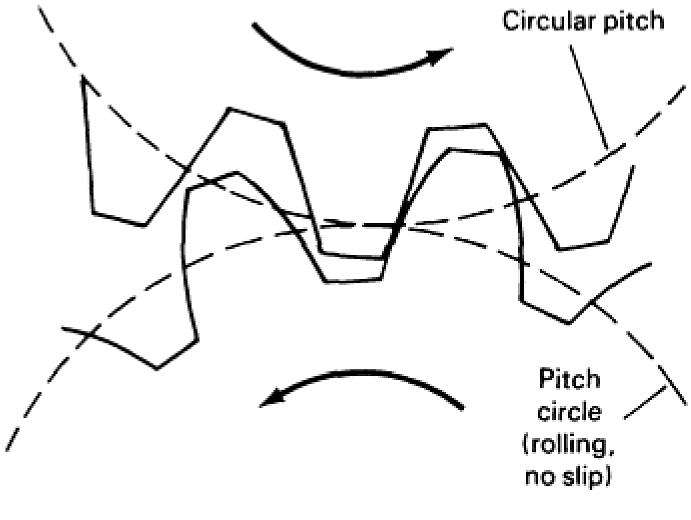


Fig. 1 Engagement of gear teeth in a gear set

Physical Signs of RCW

Gears and bearings are the components most commonly subjected to RCF and RCW. Rolling mill rolls also frequently experience RCW (see Ref 1 for a thorough discussion). Surface hardening, refinishing, or hard coating is sometimes used to attack the problem. Rolling contact wear is often manifested by defects on the rolled product that replicate the worn portions of the roller surfaces. See the article "Metalworking Lubricants" in this Volume for a discussion of lubricant selection.

Gears. As previously mentioned, the wear modes experienced by gears usually involve both rolling and slip or sliding. Dudley (Ref 2) has described the wear of gear teeth, noting that the relative amount of sliding increases as the number of teeth increases. Furthermore, the wear is not generally a function of the relative sliding velocity along a tooth face; however, at very high speeds, scoring can occur, and the friction losses in gear sets are strongly affected by sliding

velocity. Hypoid gears, worm gears, and spiroid gears can run relatively well even when appreciably worn; consequently, gear designers do not always use wear as a critical design parameter. Major considerations in gear design are that:

- Gear tooth stress does not exceed a critical value for desired life
- Gear materials are of the right kind and quality
- The form and finish are adequate
- The lubricant and lubrication system are adequate
- The system is adequately protected from rust and contaminants
- Pitting fatigue life versus contact stress and elastohydrodynamic lubrication conditions

However, determination of the size of the gear unit--that is, the pit

ch diameter, gear ratio, and the face width--often must involve consideration of the Hertzian conditions necessary to avoid pitting failure of the gear face within the design life. Several types of pitting are recognized with regard to gear surface fatigue failures:

- *Initial pitting.* Surface fatigue that usually occurs as a narrow band just below or at the pitch line at the beginning of component operation and that stops after the asperities have widened sufficiently (worn-in) to carry the load adequately
- *Destructive pitting.* Usually starts below the pitch line and spreads in number and size of pits until the gear shape is rendered unusable

Other terminology has been used for describing surface damage in gears. Terms include micropitting, surface origin pitting, subsurface origin spalling, subcase fatigue (also known as case-crushing), scuffing, and plastic flow.

Spalling is another type of wear mode that occurs in RCW of gear teeth. Spalling occurs sporadically, its frequency and exact location being statistical. The relatively large particles produced by surface spalling can cause damage if they embed in the contact surface or find their way into clearance-critical portions of gears or bearings. Dudley (Ref 2) states that both destructive pitting and spalling seldom happen at less than about 10,000 cycles, but notes that plastic flow (primarily in metallic or polymeric gears) can occur at a relatively low number of cycles if there is an overload on the contact surfaces. In summary, there generally are three types of gear wear: (1) normal or polishing wear, (2) moderate wear that is not necessarily destructive to gear life, and (3) destructive wear. For more details, see the article "Metalworking Lubricants" in this Volume.

Rolling-Element Bearings. As with gears, pitting, smearing, and spalling are important RCW manifestations in rolling-element bearings. The design and wear characteristics of rolling-element bearings are discussed in the article "Friction and Wear of Rolling-Element Bearings" in this Volume and in Ref 3. The depth of the spall tends to be related to the location of the maximum Hertzian shear stress below the surface. The fatigue spalling life of a bearing, *L,* is usually defined in terms of the first appearance of a spall and is generally based on the ratio of the equivalent dynamic load, *P,* to the load capacity, *C,* obtained from the manufacturer:

$$L = \left(\frac{C}{P}\right)^p \qquad \textbf{(Eq 1)}$$

The exponent p depends on the type of bearing. For ball bearings, $p = 3$; for roller bearings, $p = 10/3$.

In rolling bearings, the design and operating conditions significantly affect the location and rate of RCW damage. For example, a given bearing ball may experience a complex rotational path as it moves between the inner and outer bearing races. A given point on the surface of the ball may not be stressed once per revolution of the inner race, but rather, because of compound motions, may bear a Hertzian contact stress only occasionally. The center of the bearing race groove, on the other hand, is repeatedly stressed each time a ball rolls by. Therefore, the race tends to accumulate localized RCF damage cycles more rapidly than the balls.

Damage in rolling-element bearings can result from a complex combination of radial loads, axial loads, eccentric loads, thrust loads, and internal (cage/retainer) loading effects. A thorough treatment of the failure of rolling-element bearings, including not only rolling contact wear but also fracture, may be found in the article "Failures of Rolling-Element Bearings" in *Failure Analysis and Prevention*, Volume 11 of *ASM Handbook,* (Ref 4). The mechanisms of surface distress are also discussed in a collection of papers published in 1985 (Ref 5). Widner and Littmann (Ref 6) also provide a comprehensive treatment of bearing damage analysis.

Rolling Contact Fatigue Testing

Because RCF is a primary process for the production of RCW, some of the methods that investigators have developed to test the response of materials to RCF conditions similar to those they may experience in service are of interest. A particularly useful reference is *Rolling Contact Fatigue Testing of Bearing Steels* (Ref 7). In the introduction to this book, the editor states, "In building any moving machine, it is of primary concern to obtain a long endurance life of the rolling bearing. In designing a bearing life test, on the other hand, a long testing time should be avoided. The test must be accelerated so that results can be obtained within a reasonably short period of time. Unfortunately, the shorter the test becomes, the farther the simulation departs from real conditions of application."

Twenty-one papers in this collection describe testing machines, methods, and effects of processing and microstructure on RCF damage. Table 1 provides a guide to the RCF testing methods described in Ref 7 and 8 and Fig. 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, and 12 illustrate the testing arrangements listed in the table. A variety of RCF machines have been developed, ranging from one-of-a-kind research instruments to commercially manufactured bearing test rigs.

Table 1 Summary of RCF testing methods

Method	Description	Ref
NASA five-ball testing apparatus (Fig. 2)	Four lower balls, freely rotating 90° apart in a separator, simulates the kinematics of a thrust-loaded bearing; the contact angle can be varied; vibration sensor detects failure in unattended tests; low-(cryo) and high-temperature testing (to 1000 °C, or 1830 °F)	7[(a)]
Flat-washer testing apparatus (Fig. 3)	16 retained balls rolling in a circle on a flat washer with a 75 mm (3 in.) OD, 50 mm (2 in.) bore, and 6.4 mm ($\frac{1}{4}$ in.) thickness; 4.17 GPa (605 ksi) contact stress; 1500 rev/min; filtered lubricant delivery system; piezo sensor detects vibration	7[(b)]
Unisteel testing apparatus (Fig. 4)	Flat washer on retained balls; hanging dead-weight load; contact stress approximately 4.5 GPa (650 ksi); 1500 rev/min; drip feed of lubricant; vibration detection system; thermocouples monitor temperature (typically 50-60 °C, or 120-140 °F)	7[(c)]
Rolling contact testing apparatus (Fig. 5)	Two hemispherically ground, toroidal rollers loaded against a round bar; 40:1 ratio of roller diameter to bar diameter; 2.7-5.5 GPa (390-800 ksi) contact stress; 12,500 rev/min; drip-feed lubrication; velocity-vibration sensor	7[(d)]
Ball-rod testing apparatus (Federal-Mogul) (Fig. 6)	Three 12.5 mm ($\frac{1}{2}$ in.) balls loaded against a rotating 9.5 mm ($\frac{3}{8}$ in.) OD center rod; 3600 rev/min; spring load on opposing tapered retaining rings; accelerometer coupled with a shutdown device; drip-feed lubrication; stress per ball typically 6 GPa (870 ksi)	7[(e)]
Cylinder-to-ball testing apparatus (Fig. 7)	Symmetrical arrangement of two 19 mm ($\frac{3}{4}$ in.) balls rolling on a 12.5 mm ($\frac{1}{2}$ in.) OD captive cylinder; coiled-spring load through a multiplying lever; small cylinder rev/min = 22,677; splash lubrication; maximum contact stress, 5.8 GPa (840 ksi)	7[(f)]
Cylinder-to-cylinder testing apparatus (Fig. 8)	Symmetrical arrangement of two 12.5 mm ($\frac{1}{2}$ in.) cylinders on two 20 mm (0.8 in.) OD captive cylinders; coiled-spring load through a multiplying lever; small cylinder, cpm = 20,400; splash lubrication; maximum contact stress less than 4.4 GPa (640 ksi); vibration sensor terminates test	7[(g)]
Ring-on-ring testing apparatus (Fig. 9)	Crowned rings rolling on their peripheries; ring diameters of 50 and 53 mm (2 and 2.1 in.) provide "no-slip" condition, but various degrees of slip are possible by changing ring diameters; typically 2000 rev/min; contact ratio measured by electrical resistance; contact stress range typically 0.98-3.9 GPa (140-570 ksi)	7[(h)]
Various types	See review article of RCF and full-scale bearing testers	7[(i)]
Multiple bearing testing apparatus (Fig. 10)	Deep-groove ball bearing design; typically 3000 rev/min; four bearings on a single center shaft; maximum contact stress, 2.9 GPa (420 ksi); accelerometers on the outer housing monitor failure	7[(j)]
Rolling four-ball testing apparatus (Fig. 11)	Top ball drives three lower balls in a tetragonal arrangement; lower balls free to rotate in the cup; all balls 12.5 ($\frac{1}{2}$ in.) diam; upper ball spindle speed, 1500 rev/min; 5.9 kN (1325 lbf) load applied vertically	7[(k)]
High-speed four-ball	Same arrangement as above, but speeds of 15,000-20,000 rev/min; operating temperatures often exceed	7[(k)]

testing apparatus (Fig. 11)	100 °C (210 °F) (Plint machine)	
"AOL" vertical testing apparatus (not shown)	11 retained balls clamped between two flat washers; thrust load; recirculating lubricant system	7[k]
Inclined ball-on-disk testing apparatus (Fig. 12)	Spindle-held 20.5 mm (0.8 in.) ball rolling on a disk; up to 800 °C (1470 °F); ball speed up to 7200 rev/min; disk speed up to 3600 rev/min; variable slide/roll ratios; traction measurements; designed for ceramics	8

Further information can be found on the following pages of Ref 7:

(a) p 5-45,
(b) p 46-66,
(c) p 67-84,
(d) p 85-106,
(e) p 107-124,
(f) p 125-135,
(g) p 136-149,
(h) p 150-165,
(i) p 169-189,
(j) p 206-218,
(k) p 219-236

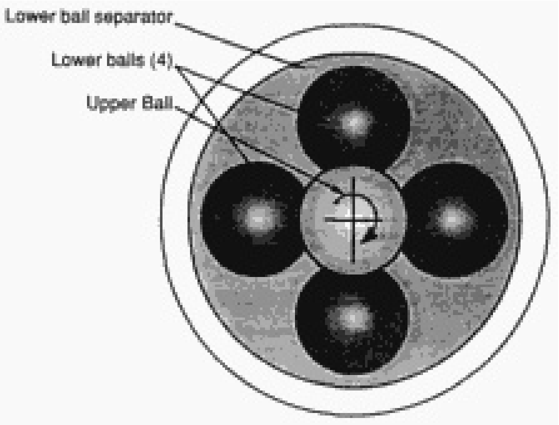


Fig. 2 NASA five-ball RCF testing apparatus. See Table 1.

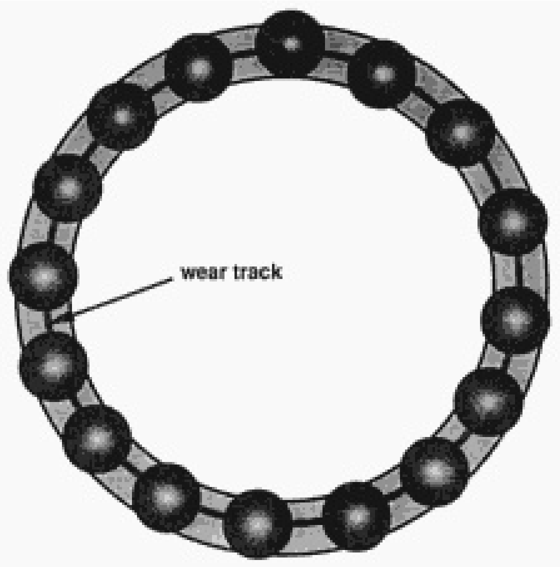


Fig. 3 Flat-washer RCF testing apparatus. See Table 1.

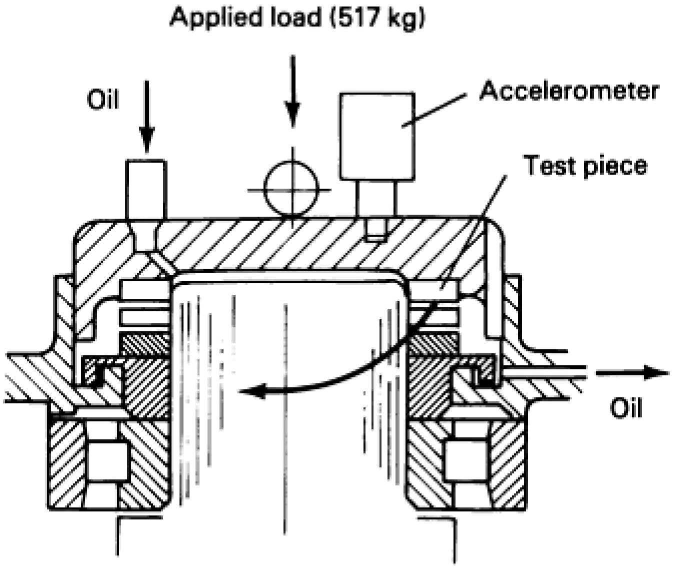


Fig. 4 Unisteel RCF testing apparatus. See Table 1.

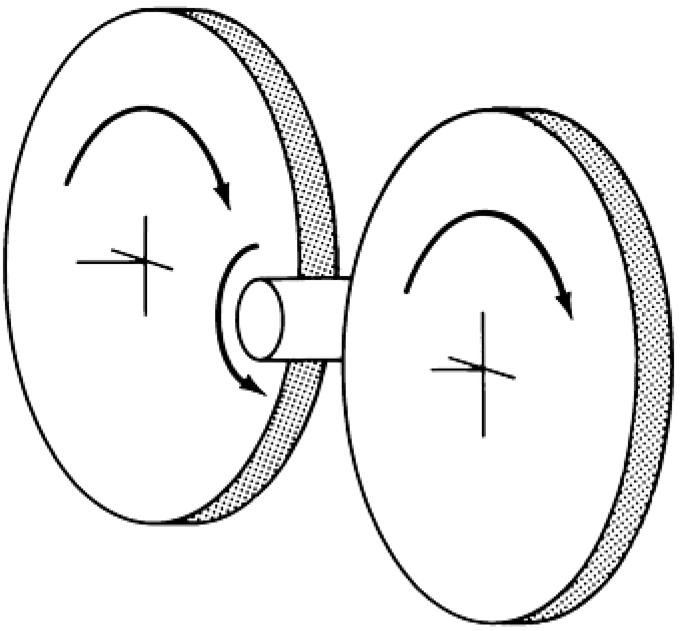


Fig. 5 Rolling contact testing apparatus. See Table 1.

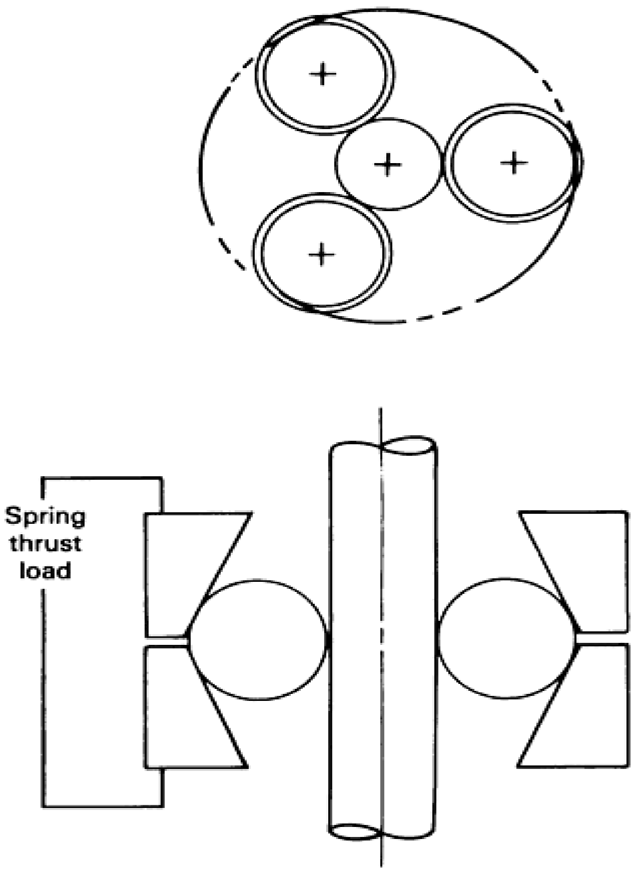


Fig. 6 Ball-rod RCF testing apparatus. See Table 1.

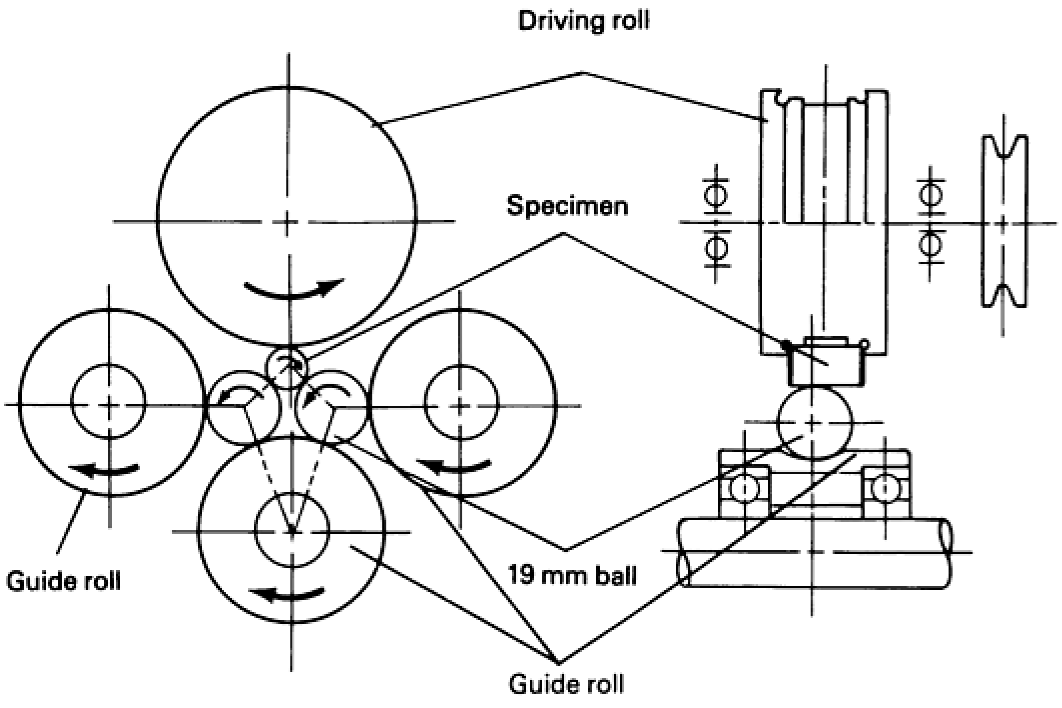


Fig. 7 Cylinder-to-ball RCF testing apparatus. See Table 1.

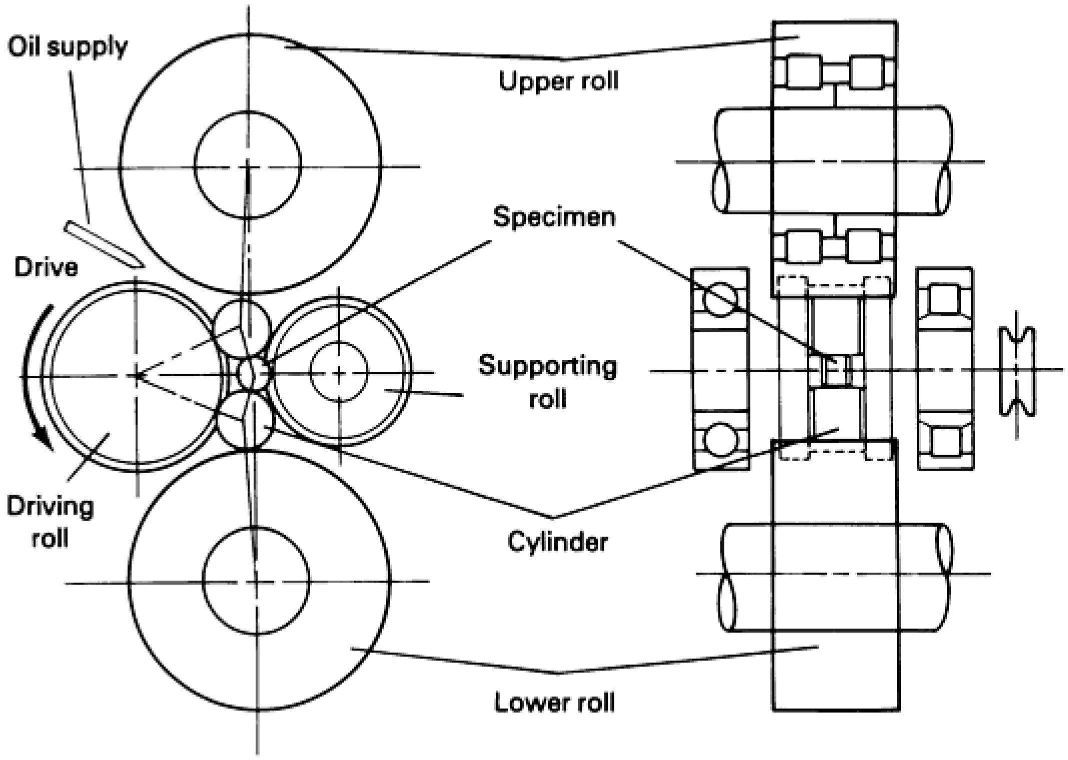


Fig. 8 Cylinder-to-cylinder RCF testing apparatus. See Table 1.

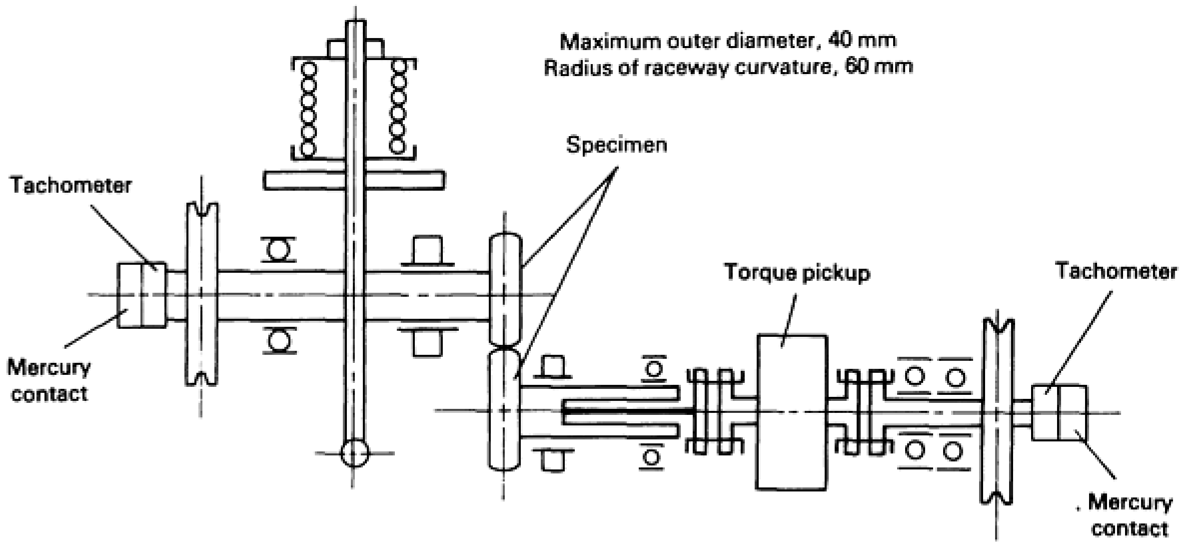


Fig. 9 Ring-on-ring RCF testing apparatus. See Table 1.

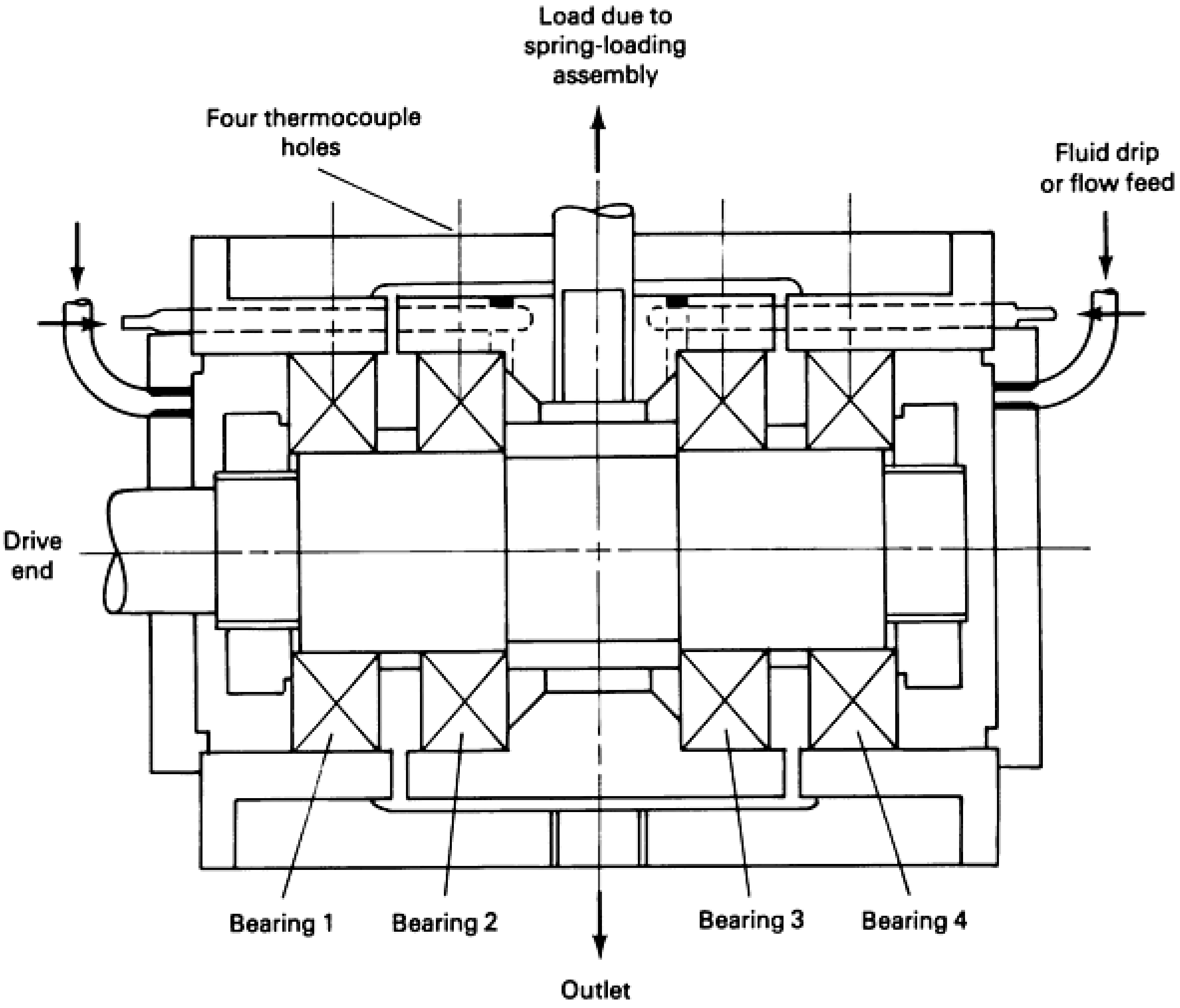


Fig. 10 Four-bearing RCF testing apparatus. See Table 1.

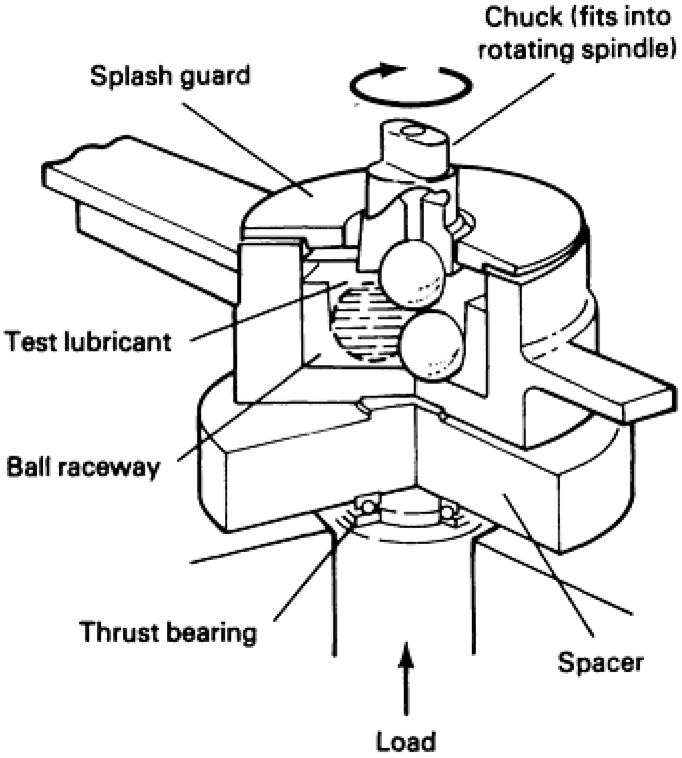


Fig. 11 Four-ball RCF testing apparatus. See Table 1.

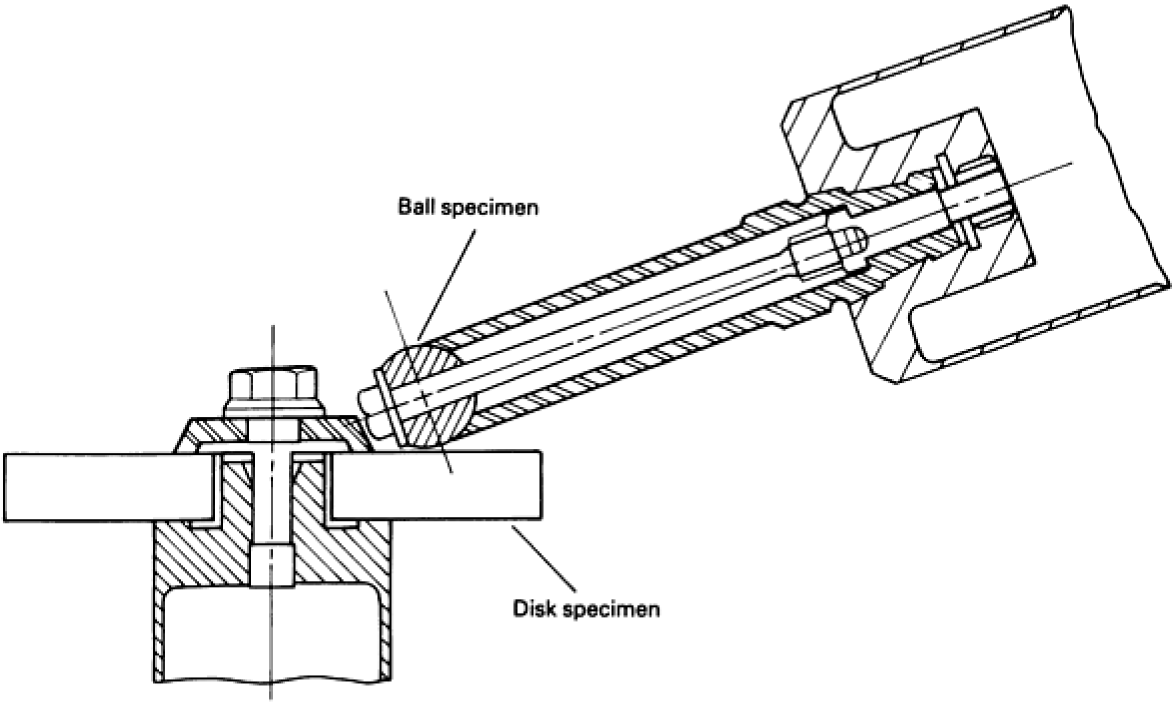


Fig. 12 Ball-on disk RCF testing apparatus. See Table 1.

Mechanisms of RCW

Dowson (Ref 9) describes some of the earliest studies of the mechanisms of bearing fatigue and wear failure performed by Goodman in the late 1800s and early 1900s. Goodman introduced the terms pitting, flaking, scratching, and peeling to the wear failure analysis vocabulary, and in a 1912 review paper introduced the concept of surface fatigue thusly: "The speed-effect on the balls has possibly some relation to the well-known effect of very rapid reversals of stress."

Hershey (Ref 10) describes the history of understanding the mechanisms of subsurface fatigue, beginning with groundbreaking work in the mid- to late 1930s that established the manner in which intersecting, subsurface fatigue cracks formed surface pits in rolling elements. It is now known that the mechanisms of rolling contact wear involve far more than subsurface fatigue. Rolling contact wear can involve mild abrasion from contaminants in lubricants, it can be affected by water or acids in lubricants, it may be exacerbated by transfer of cage material, and in addition, the contributing combination of mechanisms may change over time.

Much of the research done on the mechanisms of RCF and RCW has been based on metals and alloys; consequently, relatively little is known about the detailed behavior of ceramics and composites in similar situations. When new metal rolling components are first started up, a surface-conditioning process normally occurs. In this process, subtle changes take place in both the surface roughness and the subsurface microstructures to reach the steady-state, running condition. Initially, the surface material may be stressed beyond its elastic limit, causing some plastic deformation. As repeated rolling contact occurs, a condition may eventually be reached wherein no further plastic deformation occurs. This process is called shakedown and has been analyzed in detail by Johnson (Ref 11).

The processes of RCF and RCW generally involve the following steps:

1. Accumulation of dislocations caused by repetitive stressing of the subsurface microstructure; a dislocation cell structure may be formed
2. Nucleation of voids or microcracks in regions of maximum Hertzian stress or nearby discontinuities in the microstructure, such as grain boundaries, preexisting porosity, or inclusion/matrix interfaces
3. Propagation of the microcracks in the subsurface
4. Linking of the cracks and movement of the crack tip toward the free surface
5. Creation of flakes, pits, and/or spalls
6. Spread of the damage to adjoining portions of the surface
7. Initiation of major fatigue cracks from surface or subsurface defects, sometimes causing catastrophic fracture

A bearing or rolling component is often considered to have "failed" at the first appearance of a spall (step 5). In some systems, RCW will go undetected until step 7. As mentioned above, lubricants may accelerate step 3 by helping wedge cracks to open. The ceramic bearing industry has made a significant investment in materials development to improve the fracture toughness of their materials and to reduce the proclivity toward catastrophic failure. Composite materials are under consideration because they contain internal crack stoppers or crack-branching, energy-dissipating internal interfaces.

Associated damage processes derive from the movement of wear particles spalled from the surface of fatigued rolling elements or their counterfaces. The particles may embed in the surfaces of the rollers, score the mating surfaces, or act as third-body abrasives to alter the surface roughness and thus the lubricant flow characteristics in the interface.

The reader is referred to several reviews of RCF and RCW mechanisms. Zwirlein and Schlicht (Ref 12) discussed RCF damage mechanisms in bearing steels, including the development of "white bands" in the near-surface regions of bearings in conjunction with the decay of the martensitic structure. For 52100 bearing steel, the morphology of the white bands depended on the combination of contact pressure and number of loading cycles. Olver *et al.* (Ref 13) reviewed wear in rolling contacts with small amounts of slip and drew comparisons with Suh's delamination theory of water (Ref 14). Several differences were noted between their observations and the predictions from the delamination theory. In particular, Olver *et al.* found that cracks nucleated at or near the surface, rather than below the surface. Also, cracks tended to run obliquely to the surface in a direction opposite to that of the tractive stress, rather than parallel to the surface. Finally, the wear was dependent on the relative hardness of the contacting bodies, not on sliding distances, as stated in the delamination theory. Even with relatively small amounts of slip relative to rolling, unacceptable wear was obtained if the lambda ratio, λ(the ratio of the elastohydrodynamic lubrication film thickness to the composite surface roughness), was

small. Improved wear life can be produced by ensuring that the lubricant has adequate load-bearing capacity, in order to maintain a higher λ ratio in operating bearings.

The effect of retained austenite on the RCF of a steel containing 0.18% C, 1.5% Cr, 4.25% Ni, and 1.0% W was reported by Zhu *et al.* (Ref 15). Using carburizing followed by carbonitriding to increase the amount of retained austenite in surface layers of the steel, they conducted RCF tests with crowned 50 mm (2 in.) diam rollers. They found that the greater the retained austenite content, the greater the contact fatigue resistance, and reasoned that increased toughness from the austenite enhanced the microstructural resistance to crack propagation.

Common ferrous materials for rolling-element bearings include AISI 52100 steel, 440C stainless steel, and M-50, M-50 Nil, M-1, M-2, and M-10 tool steels. The temperature limit for 52100 and 440C is about 250 °C (480 °F), and molybdenum-containing tool steels such as M-1, M-2, and M-10 may be used up to about 500 °C (930 °F). High-reliability bearings are also made from vacuum degassed and vacuum induction melting/vacuum arc remelting (VIM/VAR) steels. Advanced silicon nitride ceramics, such as Norton Company's NBD 200, offer higher temperature capabilities and excellent RCF properties, but are considerably more expensive than steel rolling-element bearings.

A key to improving the RCF of ceramic bearings was to reduce the porosity and inclusion content of the materials so that crack-initiating microstructural features would be absent from the near-surface regions where Hertzian contact stresses are maximum. High-performance aerospace and military applications will continue to drive the development of high-temperature ceramic materials for rolling-element bearings, and commercial applications of ceramics in advanced automotive and truck engines are gradually appearing. Single-piece cam roller-followers made from silicon nitride ceramics are replacing more complicated steel parts that contain needle bearings on their inside diameter. The cost of the parts, particularly the cost of machining, will affect the rate at which they are introduced into mass-produced automobiles and trucks. The goal of producing truck engines that will run for 750,000 to 1,000,000 miles will also spur the use of ceramics, which have demonstrated outstanding reliability in engine tests.

Hybrid bearings that contain ceramic rolling elements and metallic races and retainers are also under development for military and high-performance applications. Lubrication is a major problem at elevated temperatures, and various types of solid lubricants have been developed. Solid lubrication poses the problem of ensuring a continual supply of the lubricant. See the article "Solid Lubricants" in this Volume for a more complete discussion.

The most common gear material is steel, but a variety of other materials--nonferrous, plastic, and ceramic--have also been used. The material must be able to be fabricated into the proper contour and finish for effective operation, yet possess the hardness necessary to resist wear and deformation. Generally, hardness is a major factor in the selection of gear materials, but temperature and corrosion resistance may also be important in specific applications. Surface treatments such as case hardening are used after gear profiles have been machined from softer stock. Proper selection of a compatible counterface material is also important. Table 2 provides a summary of common ferrous gear materials, their typical heat treatments, and range of surface hardnesses.

Table 2 Typical ferrous gear materials

Material	Surface treatment	Minimum surface hardness	
		HRC	**HB**
Steel			
A-1 through A-5	Through-hardened and tempered	. . .	180-400
	Flame- or induction-hardened	50-54	. . .
	Carburized and case-hardened	55-60	. . .
AISI 4140	Nitrided	48	. . .
AISI 4340	Nitrided	46	. . .
Cast iron	As-cast	. . .	175-200
Nodular iron	Annealed, quenched and tempered	. . .	140-270
Malleable iron (pearlitic)	. . .	. . .	165-240

References

1. J.A. Schey, *Tribology in Metalworking,* American Society for Metals, 1983, p 249

2. D.W. Dudley, Gear Wear, *Wear Control Handbook,* M.B. Peterson and W.O. Winer, Ed., American Society of Mechanical Engineers, 1980, p 755-830
3. L.B. Sibley, Rolling Bearings, *Wear Control Handbook,* M.B. Peterson and W.O. Winer, Ed., American Society of Mechanical Engineers, 1980, p 699-726
4. R.L. Widner, Failures of Rolling Element Bearings, *Failure Analysis and Prevention,* Vol 11, 9th ed., *Metals Handbook,* American Society for Metals, 1986, p 490-513
5. D. Dowson, C.M. Taylor, M. Godet, and D. Berthe, Ed., *Mechanisms and Surface Distress,* Butterworths, London, 1985
6. R.L. Widner and W.E. Littmann, Bearing Damage Analysis, *Proceedings of the Mechanical Failures Prevention Group,* Publ. 423, National Bureau of Standards, 1976, p 67-84
7. J.J.C. Hoo, Ed., *Rolling Contact Fatigue Testing of Bearing Steels,* STP 771, ASTM, 1982
8. L.D. Wedeven, R.A. Pallini, and C.G. Hingley, *Selection and Use of Wear Tests for Ceramics,* C.S. Yust and R.G. Bayer, Ed., STP 1010, ASTM, 1988, p 58-73
9. D. Dowson, *History of Tribology,* Longman, London, 1979, p 385
10. M.D. Hershey, *Theory and Research in Lubrication,* John Wiley & Sons, 1966, p 337
11. K.L. Johnson, *Contact Mechanics,* Cambridge University Press, 1985
12. O. Zwirlein and H. Schlicht, Rolling Contact Fatigue Mechanisms--Accelerated Testing Versus Field Performance, *Rolling Contact Fatigue Testing of Bearing Steels,* J.J.C. Hoo, Ed., STP 771, ASTM, 1982, p 358-379
13. A.V. Olver, H.A. Spikes, and P.B. McPherson, Wear in Rolling Contacts, *Proceedings of ASME Wear of Materials Conference,* American Society of Mechanical Engineers, 1985, p 254-272
14. N.P. Suh, The Delamination Theory of Wear, *Wear,* Vol 25, 1973, p 111
15. D. Zhu, F. Wang, Q. Cai, M. Zheng, and Y. Cheng, Effect of Retained Austenite on Rolling Element Fatigue and Its Mechanisms, *Proceedings of ASME Wear of Materials Conference,* American Society of Mechanical Engineers, 1985, p 572-577
16. *Mark's Standard Handbook for Mechanical Engineers,* E.A. Avallone and T. Baumeister III, Ed., 9th ed., McGraw-Hill, 1987, p 8-110 to 8-111

Impact Wear

Peter A. Engel, Department of Mechanical and Industrial Engineering, State University of New York at Binghamton

Introduction

IMPACT WEAR can be defined as the wear of a solid surface that is due to percussion, which is a repetitive exposure to dynamic contact by another solid body. This somewhat restricted definition detaches "erosive" phenomena by jets and streams of particles from the subject of study (Ref 1). Volume (or mass) removal of material, either at or under engineering contact stress levels, is discussed in this article, as opposed to permanent indentations cause by plastic deformation.

Many industries employ processes that lead to impact wear. Machine components, cams, and gears mate with a certain dynamic component. Typical applications occur in electromechanical printers; a prime example is that of typefaces, which are expected to hold definition, thus assuring high print quality, often for billions of cycles. Separable electrical connectors must make repeated contact without excessive removal of highly conductive noble or seminoble films. In drilling devices applied to a variety of media, ranging from computer boards to oil-shale, tool wear is a concern of economy. In the chattering of tubes carrying liquid wastes from nuclear reactors, safety and reliability are most important.

In most of the above applications, impact occurs with a component of sliding, compounding the relative normal approach. Thus, the term "compound impact wear" has been coined. Another consideration for the mode of contact involves the intensity, of frequency of participation, for mating surfaces. "One-body wear" is generated on one repeatedly exposed surface, as in the case of a printer typeface that is regularly impacted against continually renewed paper. However, in some devices, such as relays, the same neighboring surfaces are continually mated. If one of the mating surfaces is prone to more wear than the other, and if it is the surface of primary concern, then it could still be viewed in terms of one-body wear. If the wear of both surfaces needs to be monitored, then the situation is termed two-body wear.

The various percussive wear mechanisms include adhesive, abrasive, surface fatigue, corrosive, and thermal wear. These pure forms correspond to those that occur in sliding and rolling contacts. The fundamental differences have been sought in the nature of pressures and the friction coefficients that arise in those contact modes (Ref 2). Hybrid wear mechanisms, which combine several of the above, are frequent. An example is the fretting that occurs on the back-printer device of a hammer striking against a print band (Ref 3, 4).

The wear mechanism strongly depends on the combination of mated materials used. For example, thermal wear can occur in polymers. The contact stress range can also influence the wear mechanism. If impact speed is increased, severe adhesive wear can succeed the low-stressed fretting wear. The wear mode, such as the relative size of the sliding versus normal impacting speed component, often makes an essential difference. The worn surfaces of aluminum projectile specimens subjected to compound impact cycles, upon repeatedly impacting a steel plate rotating at different speeds, are shown in Fig. 1.

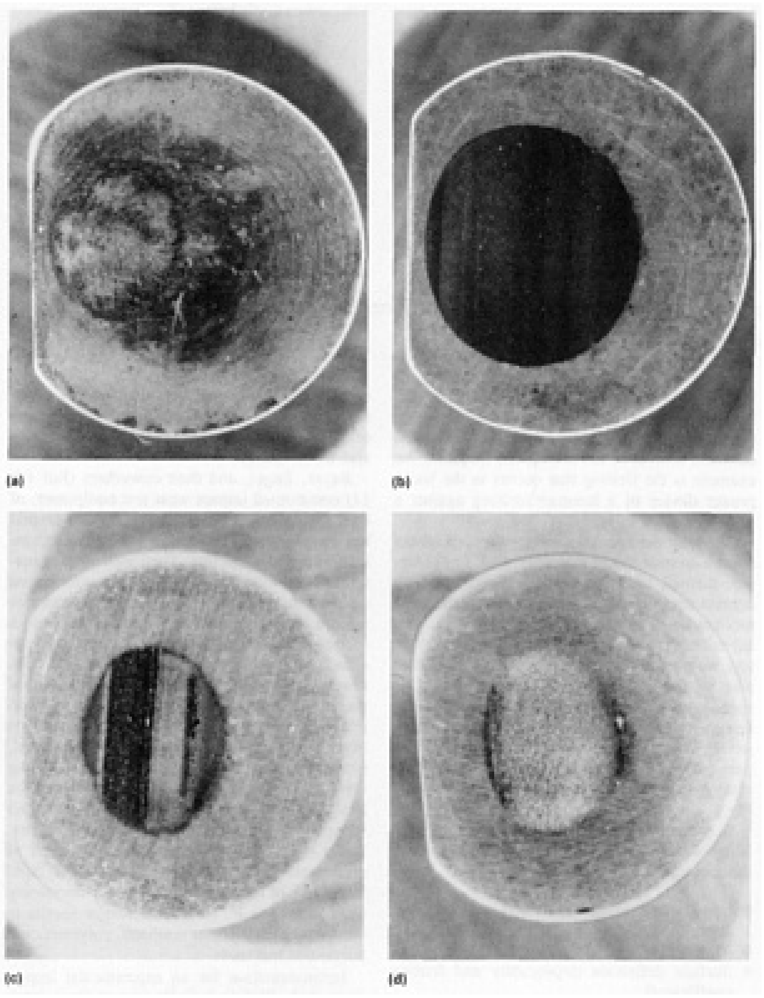


Fig. 1 Impact-worn surfaces of spherical-headed aluminum 2024-T4 projectiles, 15×; the 0.5 g (0.018 oz) projectiles (for which δ = 1.1 μm, or 44 μin., and *R* = 41 mm, or 1.6 in.) repetitively impacted an alloy steel target plate at *v* = 1.7 m/s (68 in./s), $q/\sigma_0 \approx 0.8$. (a) *v* = 0, *N* = 1.5 × 10^7. (b) *v* = 250 mm/s (10 in./s), *N* = 1.5 × 10^7. (c) *v* = 1.25 m/s (50 in./s), *N* = 10^4. (d) *v* = 3.0 m/s (120 in./s), *N* = 10^3

In order to systematically study the subject, it is helpful to divide variable parameters into these categories:

- Loads (for example, mass and speed components)
- Materials (elastic and strength data)
- Surface definition (topography and friction coefficient)
- Wear coefficients

A rational, semi-empirical impact wear theory (Ref 1), which embodies all the above ingredients, is outlined in this article.

Experimental Background

From a long line of experimental impact wear studies, several that underlie the variety of industrial needs are described below. Examining vibrational contact for nuclear engineering materials, De Gee *et al.* (Ref 5) wore sintered aluminum powder (SAP) hemispheres against SAP planes, and Ko (Ref 6) investigated the chattering of heat-exchange tubes against annular supports. Wellinger and Breckel (Ref 7) studied the wear of the plastically deforming contacts of a great variety of metal specimens subjected to repeated normal impact. Montgomery (Ref 8) correlated and modeled wear that was caused by rock drill bits. The work of Sorokin (Ref 9), one of several Soviet industrial wear investigators, is also worth citing.

Bayer, Engel, and their coworkers (Ref 10, 11) constructed impact wear test equipment, of both the ballistic projectile (Fig. 2a) and pivotal hammering type (Fig. 2b), out of electromechanical printer mechanisms. They used various material combinations, stress range, lubrication conditions, and other parameters to optimize various machine contacts. Rice (Ref 12) built a reciprocating impact wear tester and investigated, with his coworkers, a range of wear relationships (Ref 13, 14) that were dependent on metallurgy, such as grain structure. Sugita *et al.* (Ref 15) dealt with effects of impact wear on MgO single crystals, a typical brittle material. Fretting wear that was due to repeated impacting was studied by Levy and Morri (Ref 16). Sare (Ref 17) investigated wear-resistant materials subjected to abrasion in ore-crushing hammers. Bayer (Ref 18) found that the impact wear of elastomers typically exhibits two stages: a rapid initial are of deformation, with ever-decreasing rates following it. He found similar results in thermoplastic polyester urethane, polyester, and polyamide materials.

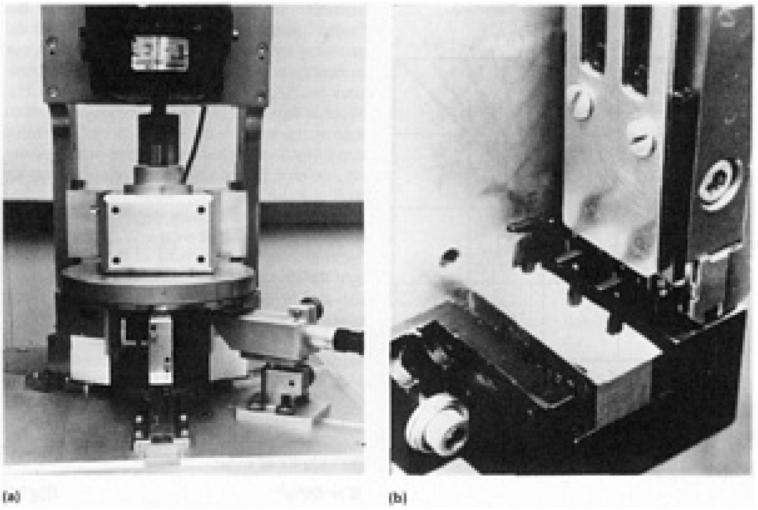


Fig. 2 Impact wear testing apparatuses. (a) Ballistic impact wear tester with three projectile bays located 120° apart. (b) Pivotal hammering tester, where three hammers impact a polymer sheet

Instrumentation for an experimental impact wear study (Ref 1) includes impact force measurement devices (such as piezoelectric transducers) that calibrate force versus speed relations. The contact stress for either a projectile or hammer that strikes a known material at known speed can be calculated by analytical or numerical (finite-element) methods. This

information, which is related to the Hertz contact theory, turns out to be important in terms of the wear theory outlined in this article.

The wear scar can be measured by removing the worn specimen and topographically scanning it. It can then be repositioned for more load cycles if an accurate method is available (Ref 19). Scanning electron microscopy (Ref 20) is commonly used to check the wear mechanism, the wear scar shape, and any idiosyncrasies of the process. The worn surface must be characterized for hardness (for example, by a Rockwell, Vickers, or Knoop microhardness test) (Ref 1, 21).

Model for Compound Impact

To model compound impact (Ref 22), Fig. 3 shows a projectile of mass, *m*, approaching a target plane with a normal relative speed, *V*. Subsequent to striking (at time, *t*, equated to 0) the plane that is moving in the tangential *x*-direction at a speed, *v*, the projectile is accelerated by friction in the *x*-direction. Because friction is impending, the equation of motion of the slipping projectile includes the Coulomb friction term:

$$\frac{d\dot{x}}{dt} = \mu P(t),\ t > 0 \quad \textbf{(Eq 1)}$$

with the initial conditions

$$t = 0,\ x = 0,\ \dot{x} = 0 \quad \textbf{(Eq 2)}$$

Approximating the impact force pulse as a sinusoidal of peak force, P_0, and duration, t^*:

$$P(t) \approx P_0 \sin\left(\frac{\pi t}{t^*}\right) \quad \textbf{(Eq 3)}$$

integration of Eq 1 yields the speed to which the mass is accelerated by time, *t:*

$$\dot{x} = \frac{\mu P_0 t^*}{\pi m}\left(1 - \cos\frac{\pi t}{t^*}\right) \quad \textbf{(Eq 4)}$$

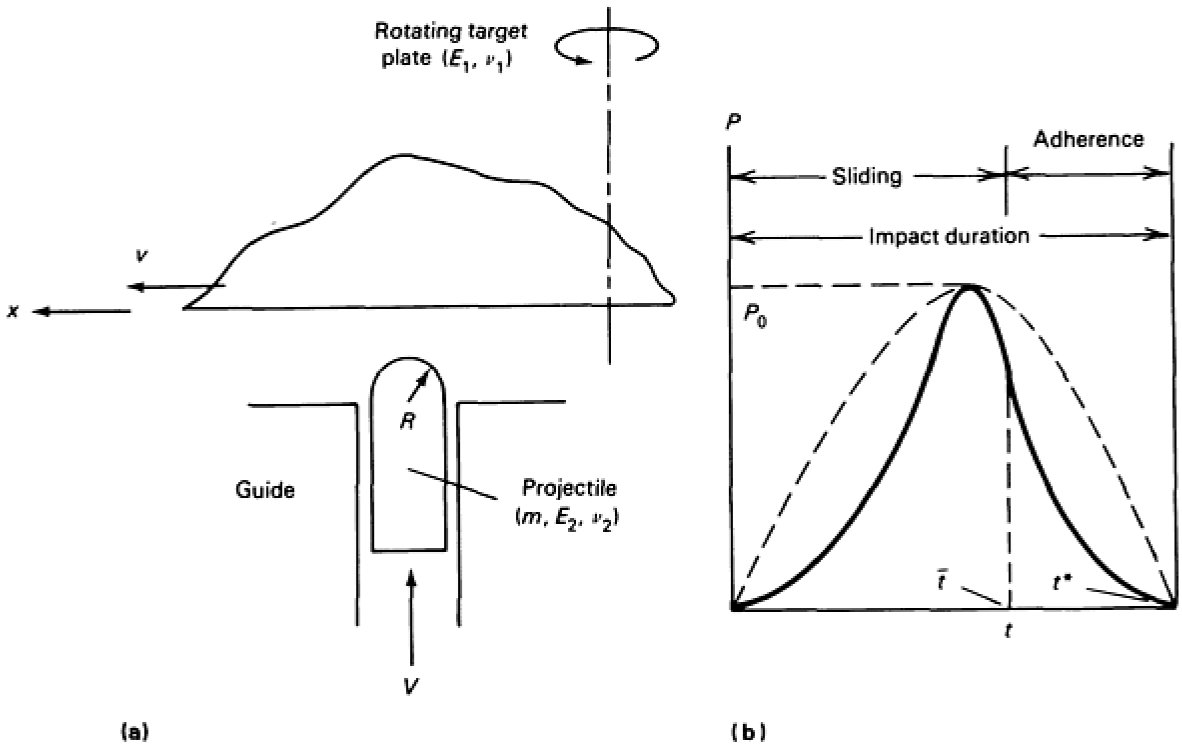


Fig. 3 Model of compound impact, where projectile strikes a tangentially moving plate. (a) Projectile approach. (b) Impact pulse and sinusoidal approximation

If the impacting mass is to catch up with the plane at some time, $\bar{t} < t^*$, then the slipping time, $\bar{t}$, can be calculated from Eq 4, by setting $\dot{x} = v$ at $t = \bar{t}$, obtaining:

$$\bar{t} = \frac{t^*}{\pi} \cos^{-1}\left(1 - \frac{\pi m v}{\mu P_0 t^*}\right) \qquad \textbf{(Eq 5)}$$

It is understood that after $t = \bar{t}$, the projectile is adhered to the plane and no longer slips against it. The negative component of the arc cos function is characteristic of the slipping properties of the system, and is called the slip factor,

$$f = \frac{\pi m v}{\mu P_0 t^*} \qquad \textbf{(Eq 6)}$$

the numerical ranges of which are also important for wear considerations:

- $f = 0$ is normal impact
- $0 < f < 2$ is moderate sliding speed; slipping gives way to adherence at time $\bar{t}$ during the impact; $\bar{t}$ is real valued in Eq 5, $\bar{t} < t^*$
- $f > 2$ is high sliding speed; slipping persists throughout the impact contact time, and is imaginary in Eq 5

If a spring, *k*, is restraining the mass in the *x*-direction (Fig. 4a), then a force $k \cdot x$ will act on the projectile, tending to increase the slipping time. The latter is calculated from a transcendental equation corresponding to Eq 5:

$$\cos\omega\bar{t} - \cos\phi\bar{t} - f\left(1 - \frac{\omega^2}{-\phi^2} = 0\right) \quad \textbf{(Eq 7)}$$

where $\omega = \sqrt{k/m}$ and $\phi = \pi/t^*$. Figure 4(b) shows the effect of spring restraint.

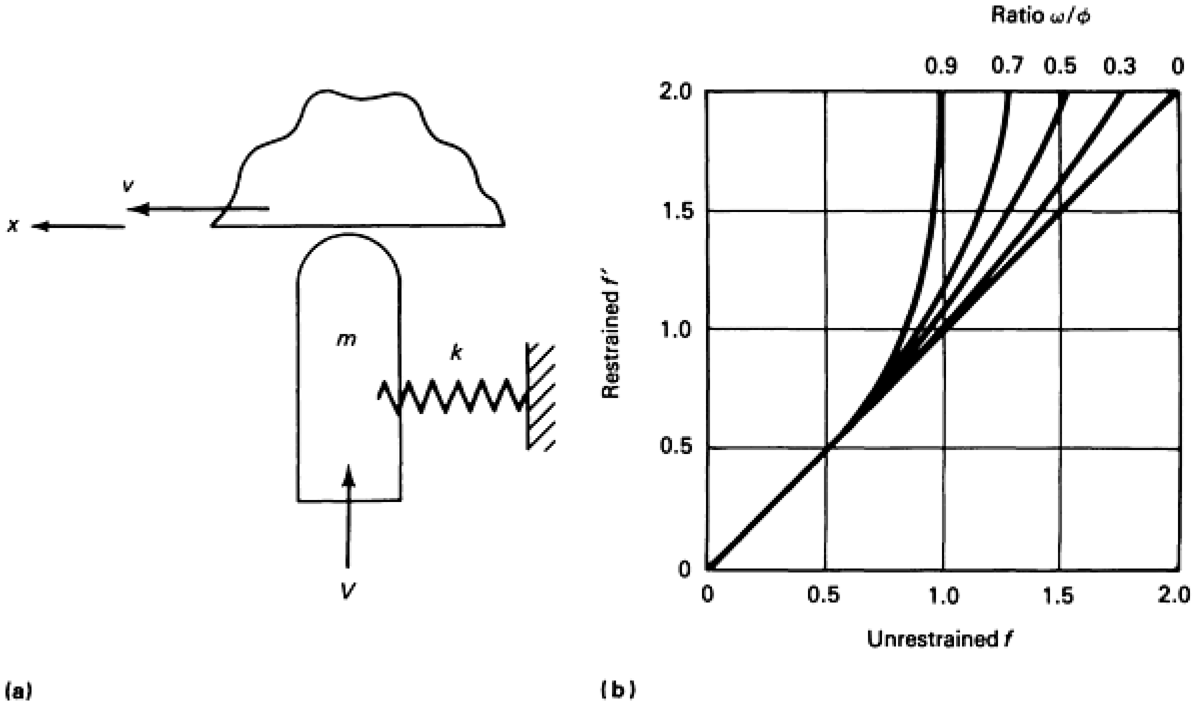


Fig. 4 Spring-restrained projectile model. (a) Impact. (b) Corrected slip factors for restrained projectiles

If rotation of the impacting projectile is considered in addition (Fig. 5), then an inertial couple, $I\ddot{\theta}$, must be added to the free-body diagram of the projectile. The slipping time is obviously shortened by this additional degree of freedom, because adherence to the target is easier to achieve. Now $\bar{t}$ is calculated from the equation

$$\frac{\cos\omega\bar{t} - \cos\phi\bar{t}}{1 - (\omega^2/\phi^2)} + \frac{mL^2}{I_{cg}}(1 - \cos\phi\bar{t}) = f \quad \textbf{(Eq 8)}$$

which corresponds to Eq 5 of the unrestrained, unrotating case of the impacting projectile. The sliding restraint and rotational capacity can be used to more accurately model pivotal hammering devices.

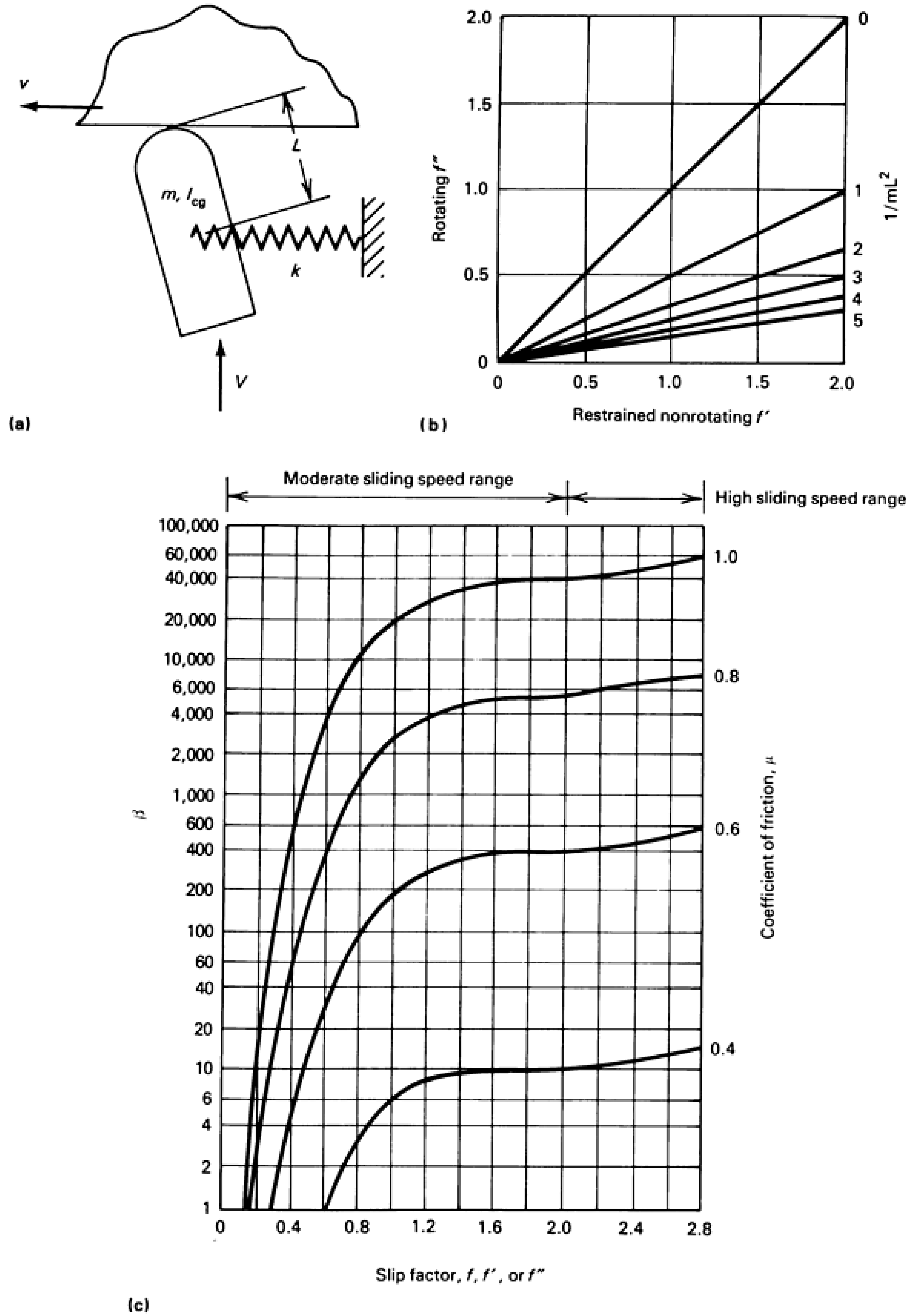


Fig. 5 Spring-restrained projectile rotating during impact. (a) Impact schematic. (b) Corrected slip factors for restrained rotating projectiles. (c) Plots of β versus *f* and μ

Linear Impact Wear

Two cases were found to be of great practical importance in repetitively impacted contacts. Discussed first is a linear wear mechanism that occurs in print heads.

Consider the linear wear mechanism that is characteristic of an abrasive sliding process:

$$W = \frac{KPX}{H} \quad \textbf{(Eq 9)}$$

where K is the wear constant for the material pair, P is the normal contact force, X is the sliding distance, H is the hardness of the wearing body, and W is the volume worn away. When a contact, such as a print head that strikes, paper, undergoes a cycle of compound impact, the one-body wear of the print head per differential time, dt, during a single impact cycle, at time, $t < \bar{t}$, can be evaluated. After adherence takes place, no more wear generation is assumed (Ref 1). Thus,

$$dW = KP(t)dX(t) \quad \textbf{(Eq 10)}$$

where x is the print head displacement, and X is the slip of the print head with respect to paper:

$$X = vt - x \quad \textbf{(Eq 11)}$$

having nonzero values in the slip region $0 < t < t < t^*$ (see Eq 5). Differentiating Eq 11,

$$dX = (v - \dot{x})dt \quad \textbf{(Eq 12)}$$

Integrating over the time $0 < t < \bar{t}$, the wear per impact cycle becomes, for the moderate sliding speed range ($f < 2$), by integration of Eq 10:

$$\Delta W = \frac{Kvt^*P_0}{\pi H} C(f) \quad \textbf{(Eq 13)}$$

$$C(f) = \frac{f}{2}, f < 2 \quad \textbf{(Eq 14)}$$

Likewise, for the large sliding speed range ($f > 2$), integration over the whole pulse yields Eq 13 with

$$C(f) = 2\left(1 - \frac{1}{f}\right), f > 2 \quad \textbf{(Eq 15)}$$

Thus, $C(f)$ is a two-part, but continuous, function converging to $C(f) = 2$ at $f = \infty$(Fig. 6).

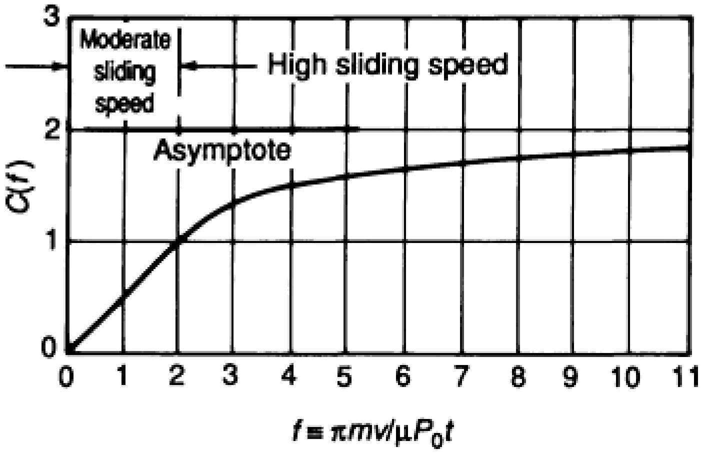


Fig. 6 Function *C*(*f*) relating abrasive impact wear to slip factor

The above wear process is termed "linear," because the wear per pulse, ΔW, is independent of the number of cycles, or of the geometry of the wear scar as consecutive impacts are incurred. A similar analysis was also applicable to computer print wires, which were observed to wear in a "pencil-sharpener" mode (Ref 1).

Abrasive wear constants, *K*, were taken from the wheel tester invented by Roshon (Ref 23), and adopted as an ASTM standard to grade computer paper for abrasivity. A wide ribbon is stretched on the cylindrical surface of a ***d*** = 1.2 m (4 ft) diameter steel drum, against which a cantilevered bob made of the print head material of known hardness, *H,* is squeezed, producing a controllable normal force, *P*, on the contact. The bob is given a slight tangential speed perpendicular to the sliding direction, so that the paper surface tends to be renewed with subsequent wheel rotations. Wear volume measurements at the contact point of the bob after a known number, *n,* of wheel revolutions at constant speed finally yield the wear constant from Eq 9:

$$K = \frac{HW}{\pi d n P} \qquad \textbf{(Eq 16)}$$

Impact Wear of Machine Contacts

The repetitive impacts that take place in metallic machine contacts tend to create a wear scar, which appears after a certain incubation period in the wear life. This means that within a period of initial cycles, N_0, called the zero-wear limit, the contact spot maintains the original surface median or root mean square (rms) height, δ.

Figure 7 shows a set of wear curves, ***h*** versus *N*, for carbon steel ballistic impact wear projectiles (*m* = 1.27 g, or 0.05 oz, Rockwell hardness (HRC) = 20, *R* = 140 mm, or 5.5 in., head radius) impacted against an alloy steel plate at the common normal approach speed, *V* = 1.73 m/s (68 in./s). The target disk rotation was varied, however, producing the relative sliding speeds, *v* = 0, 0.25 m/s (10 in./s), 1.27 m/s (50 in.s/s), 3.81 m/s (150 in./s), and 7.62 m/s (300 in./s).

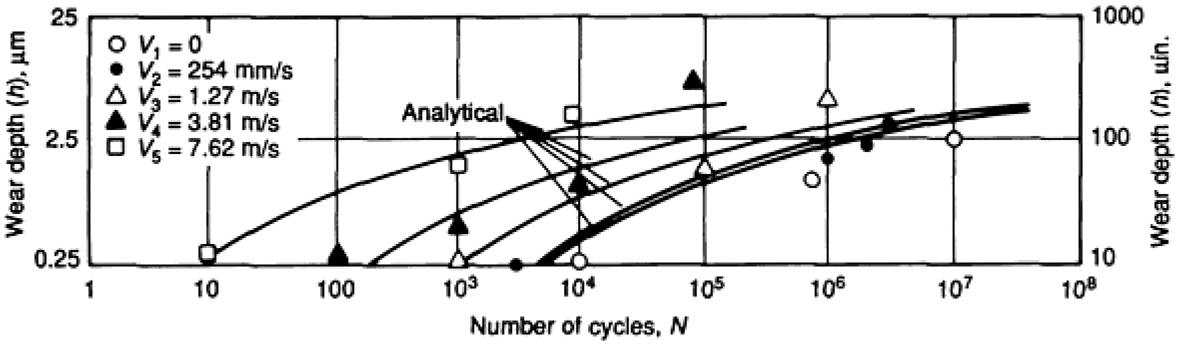


Fig. 7 Compound impact wear of carbon-steel projectiles tested in ballistic impact wear apparatus. Projectiles are carbon steel, HRC = 20, V8 ($\delta \simeq 0.5$ μm, or 20 μin.), *m* = 1.27 g (0.045 oz), *R* = 140 mm (5.5 in.), *V* = 1.7 m/s (68 in./s); target disk is alloy steel 4140, HRC = 40 to 45, V16, μ= 0.62

Projectiles were inspected for wear at logarithmic intervals. It was observed from the experimental measurement points that the wear process became "measurable" at earlier cycle numbers as the *v*/*V* ratio was increased, that is, the sliding contribution was increased with respect to the normal impact contribution.

Zero-Wear Limit. In order to establish the zero-wear limit as a quantitative, semi-empirical relation, an extension of the Bayer-Ku zero-sliding-wear theory (Ref 24) was made. The latter has, for the number of passes, *N*, denoting zero wear:

$$N_0 \left(\frac{\tau_{max}}{\gamma_r \tau_y} \right)^9 = 2000 \qquad \text{(Eq 17)}$$

In impact wear, it is sensible to consider subsurface damage, D_1, which is caused by the maximum shear stress that occurs in the depth of a contact without any shear tractions attending, and surface damage, D_2, which is due to shear traction caused by superposed sliding effects that occur during the slipping time of compound impact cycles.

The damage contributions (Ref 1) are defined as integrals of the representative maximum subsurface and surface stresses over an appropriate interval of the impact pulse. Thus,

$$D_1 = \int_0^{t^*} \tau_{max}^9 (t) dt \qquad \text{(Eq 18)}$$

is integrated over the whole pulse, but

$$D_2 = \int_0^{\bar{t}} [\mu \sigma_{z,max}(t)]^9 \, dt \qquad \text{(Eq 19)}$$

is integrated only over the slipping time, $\bar{t}$. Note that in evaluating the damage factors D_1 and D_2, the sinusoidal impact force pulse of Eq 3 can be used for a simpler calculation.

Defining the surface damage contribution factor

$$\beta \equiv \frac{D_2}{D_1} \qquad \text{(Eq 20)}$$

the zero-wear law is now formulated analogously to Eq 17:

$$N_0 \left(\frac{\sigma_{z,\max}}{\gamma \sigma_y} \right)^9 = 2000 \quad \text{(Eq 21)}$$

which has a coefficient γ that is to be determined via experiments.

From many sets of experiments conducted on various steel and aluminum projectiles, γ was found to be close to 1.1. Note that an essentially constant γ_r is another feature of the Bayer-Ku sliding wear theory. The relationship of these two constants is discussed in Ref 1.

Approximate expressions (Ref 25) of β for common metals (Fig. 5c) are:

$$\beta = \left(\frac{\mu}{0.31} \right)^9 (1.23) \int_0^{\cos^{-1}(1-f)} \sin^9 \zeta \, d\zeta, f < 2 \quad \text{(Eq 22)}$$

$$\beta = \frac{f}{2} \left(\frac{\mu}{0.31} \right)^9, f > 2 \quad \text{(Eq 23)}$$

Measurable Wear. Past the zero-wear limit, a wear scar becomes evident, and its dimensions undergo a growth process. Only symmetric quadratic (spherical or cylindrical) contact surfaces subjected to repetitive impacts are considered here. It is assumed that throughout the wear history, either the impact momentum or peak impact force remains constant. However, it is recognized that because of changes in contact geometry that are due to wear, the contact area and, thus, the Hertz contact stress state per impact, will vary. Figure 8 illustrates the descriptors relevant to measurable wear.

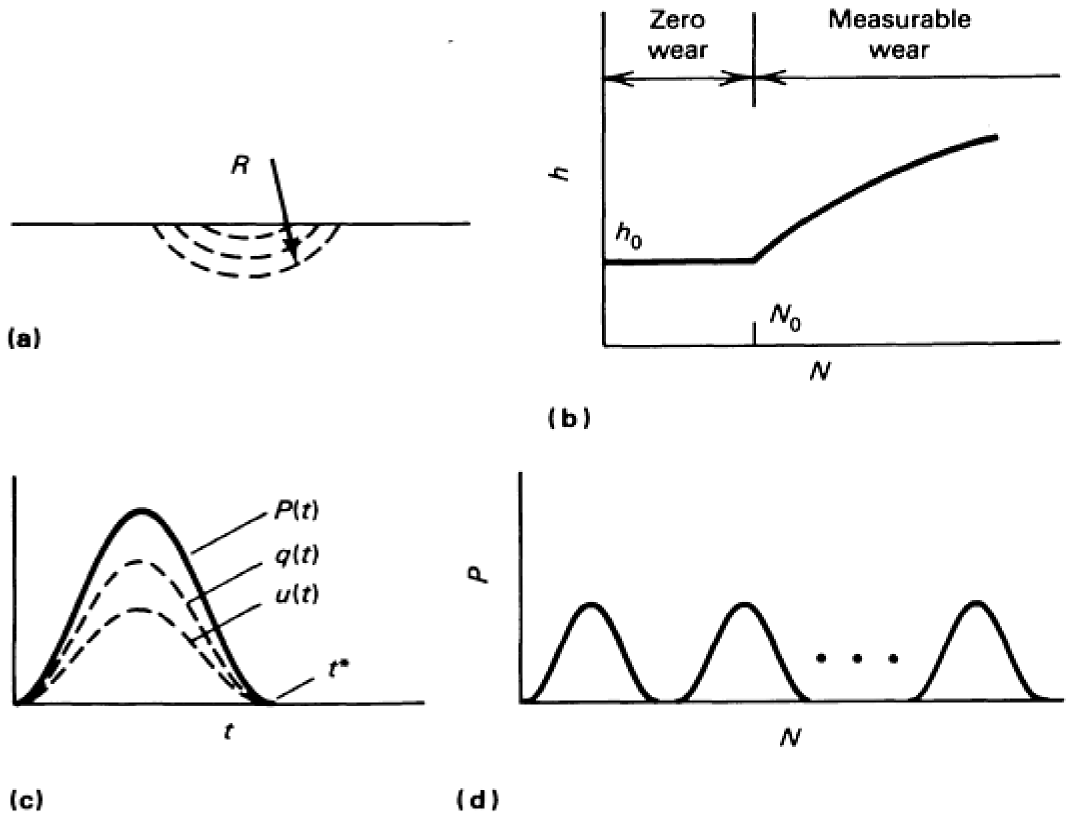


Fig. 8 Concepts relevant to measurable wear. (a) Wear formation. (b) Wear history. (c) Impact history (pulse). (d) Loading history

It is useful to establish the concept of the instantaneous wear state, which includes the geometry of the worn contact, and the Hertzian contact stresses arising in this contact, which are due to the prescribed impact cycles. The instantaneous wear state at the peak of the *N*th impact would include the instantaneous peak contact radius, *a*, peak maximum pressure, *q*, and others.

The common experience with moderately stressed machine contacts is the tendency of wear scars to evolve toward conformance of the two mating surfaces. Figure 9 shows an originally plane spring steel target gradually conforming to the outline of a tool steel pivotal hammer. In one-body wear processes, the "softer" partner progressively wears toward the geometry of the nonwearing, "harder" partner (a greater resistance to wear is meant here by hardness because it is usually proportional to *H*). Wear is dependent on the instantaneous contact area at any stage of the wear history. The initial condition to this phase is the zero-wear geometry, defined by the quadratic surface belonging to $N = N_0$, $h = \delta/2$.

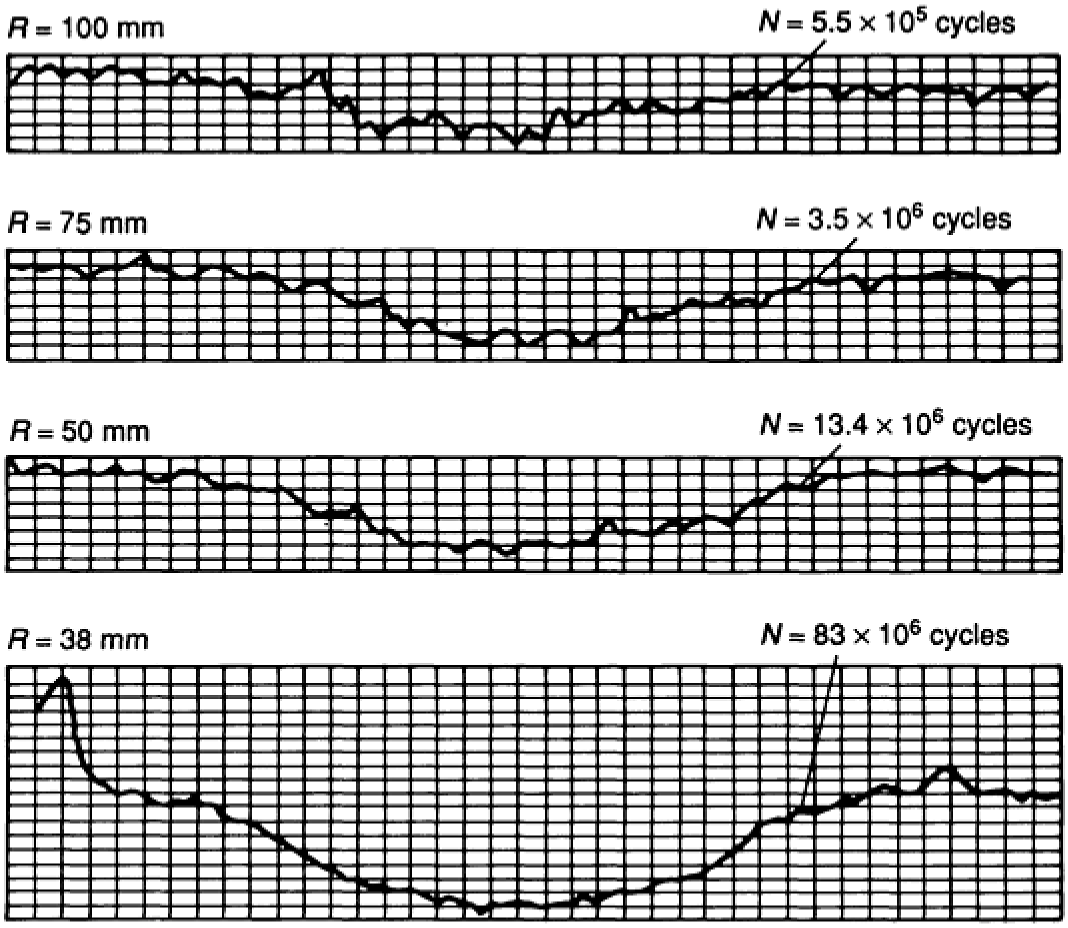


Fig. 9 Process of changing curvature on flat blued spring steel target (AISI-C1095, 48 TO 51 HRC) surface impacted by hard cylindrical hammer (Carpenter extra, air-hardened tool steel, HRC = 62 to 64, *r* = 35 mm, or 1.38 in.). In terms of scale, each vertical division is 0.508 μm (20 μin.) and each small horizontal division is 50.8 μm (0.002 in.).

In the measurable wear stage of the wear history, the growth of the wear scar can be followed by observing the deepening of the wear scar, *h*, and the change of the radius of curvature, *R*. It can also be described by the instantaneous peak contact radius, *a*, or peak contact pressure, *q*, because *a* and *q* are uniquely related by Hertz theory to the shape of the contact area defined by *h*. It is also argued (Ref 1) that *a* coincides with the instantaneous radius of the wear crater.

Let the mathematical expression of the wear mechanism be written in the following form for the volume of wear:

$$W = KN^b q^{3c} \quad \textbf{(Eq 24)}$$

The differential form of the above, in terms of both the number of cycles and the instantaneous peak pressure:

$$dW = \frac{\partial W}{\partial N} dN + \frac{\partial W}{\partial q} dq \quad \textbf{(Eq 25)}$$

Letting $b = 1$, and inserting Eq 25 into Eq 24, the differential relationship becomes:

$$dW = \frac{W}{N} dN + \left(\frac{3cW}{q}\right) dq \quad \textbf{(Eq 26)}$$

The above equation can then be integrated for particular geometries characterizing the percussive wear. Two such simple geometries are a soft ball wearing against a hard plane and a soft plane wearing against a hard ball. In the methods of solution described below, the object is to solve for nondimensional wearing geometries. These can be converted into physical geometric parameters through multiplication by the appropriate constants.

Solution Methods for Measurable Wear

Example 1: Soft Ball Worn by Hard Plane.

The ball wears from the original radius R_1 to a radius $R > R_1$. With reference to Fig. 10, the wear is the difference between two paraboloids

$$W = \pi[R_1(h_1 + h)^2 - Rh_1^2] \quad \textbf{(Eq 27)}$$

with

$$h_1 = \frac{a^2}{2R_1} \quad \textbf{(Eq 28)}$$

where the wear scar radius a was equated with the edge of the peak contact radius at the stage, N, of wear history. For the depth of wear from this geometry:

$$h = \frac{a^2\left(1 - \frac{R_1}{R}\right)}{2R_1} \quad \textbf{(Eq 29)}$$

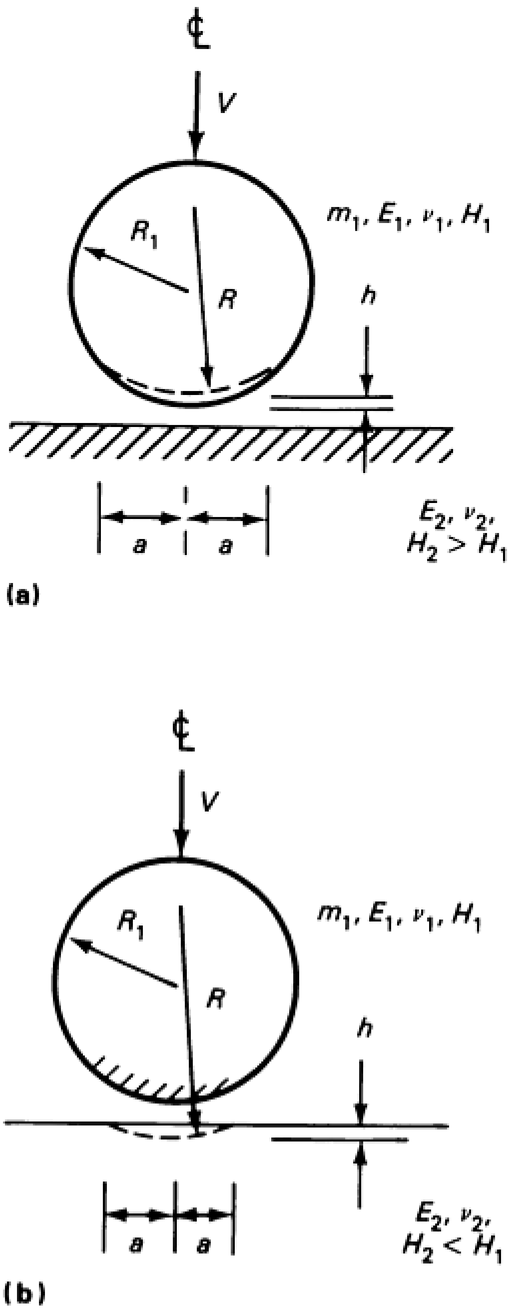


Fig. 10 Percussive wear schemes. (a) Wear of soft ball versus hard plane. (b) Wear of soft plane versus hard ball

Introducing the nondimensional curvature of the wear scar:

$$\lambda \equiv \frac{R_1}{R} \qquad \textbf{(Eq 30)}$$

Hertz contact analysis yields the following expressions for *q*, *a*, and *W* in terms of λ:

$$q = K_1\lambda^{3/5} \qquad \textbf{(Eq 31)}$$

$$a = K_2\lambda^{-2/5} \qquad \textbf{(Eq 32)}$$

$$W = \frac{\pi K_2^4}{4R_1}(1 - \lambda)\lambda^{-8/5} \qquad \textbf{(Eq 33)}$$

where elements of the Hertz impulse

$$P_0 = 0.811(V^6 m^3 R_1 E_r^2 \lambda^{-1})^{1/5} \qquad \textbf{(Eq 34)}$$

$$t^* = 4.525(V^{-1} m^2 R_1^{-1} E_r^{-2} \lambda)^{1/5} \qquad \textbf{(Eq 35)}$$

were used to obtain the constants

$$K_1 = 0.1258(V^2 m R_1^{-3} E_r^4)^{1/5} \qquad \textbf{(Eq 36)}$$

$$K_2 = 1.2411(V^2 m R_1^2 E_r^{-1})^{1/5} \qquad \textbf{(Eq 37)}$$

Substituting into Eq 26, the ordinary differential equation for wear is obtained:

$$\frac{(3 + 9c)\lambda - (8 + 9c)}{5\lambda(1 - \lambda)}\,d\lambda = \frac{dN}{N} \qquad \textbf{(Eq 38)}$$

For a value of $c = 3$, found to be practical, there results:

$$\frac{6\lambda - 7}{\lambda(1 - \lambda)}\,d\lambda = \frac{dN}{N} \qquad \textbf{(Eq 39)}$$

The wear depth, *h*, is related to λ as

$$h = C\lambda^{-4/5}(1 - \lambda) \qquad \textbf{(Eq 40)}$$

where the factor *C* has the dimension of displacement, and connotes the severity of the impact:

$$c = 0.77(V^4 m^2 R_1^{-1} E_r^{-2})^{1/5} \qquad \textbf{(Eq 41)}$$

In order to express the initial condition at the zero-wear limit, Eq 40 can be solved for $\lambda = \lambda_0$. Figure 11 facilitates the conversion from dimensional to nondimensional variables, *h* to λ, and vice versa. Note that Fig. 11 also includes cylindrical geometries, for which the wear equations can be found in Ref 1.

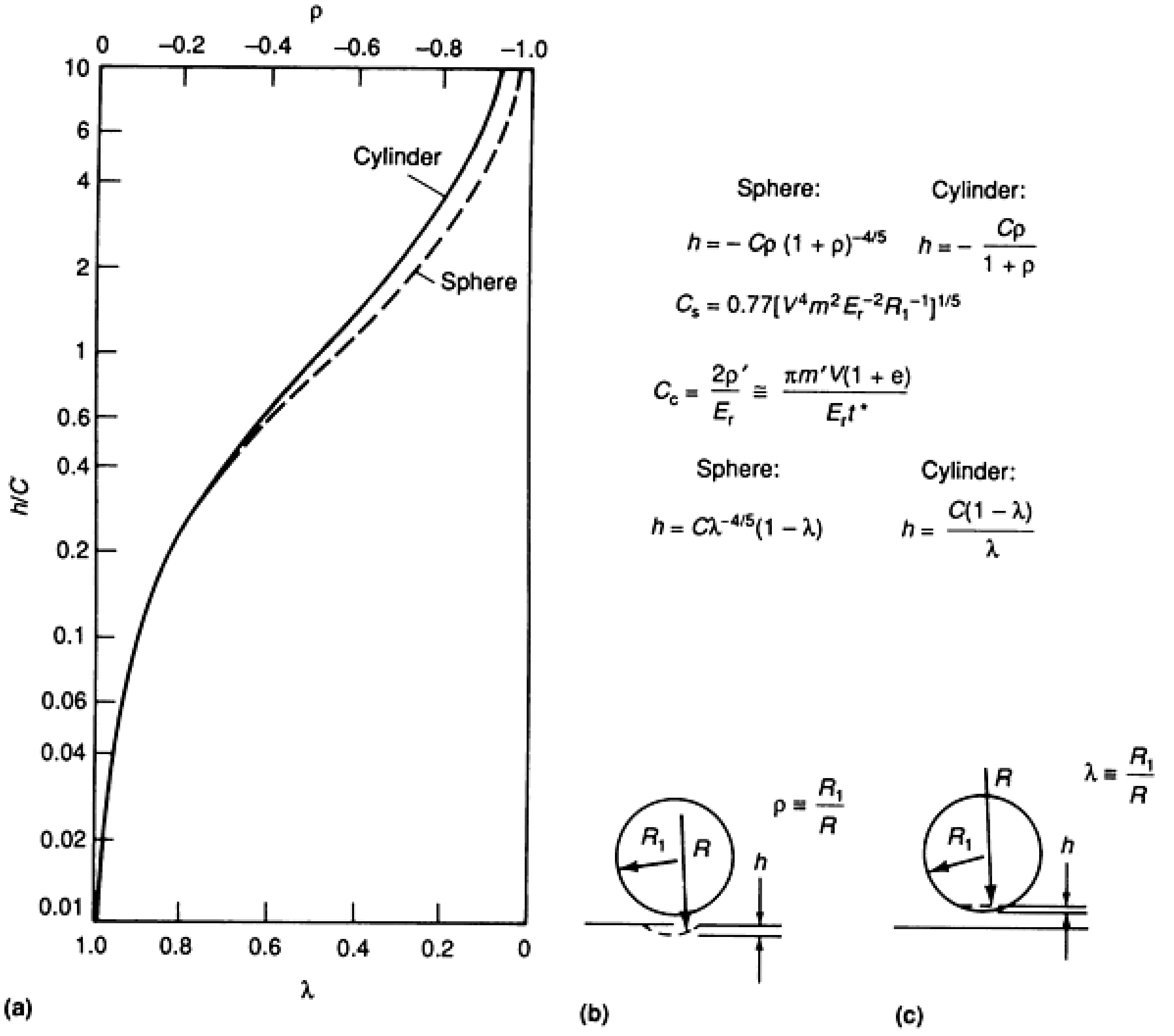


Fig. 11 (a)*h*/*C* variation versus λ and ρ for spherical and cylindrical contacts. (b) Hard sphere or cylinder wearing down soft plane. (c) Soft sphere or cylinder wearing against hard plane

Example 2: Soft Plane Worn by Hard Ball.

Figure 10(b) shows the geometry of the deepening wear scar of the plane, attending a continuously sharpening negative radius, *R.* The nondimensional curvature is taken as

$$\rho \equiv \frac{R_1}{R} \qquad \textbf{(Eq 42)}$$

All of the mathematical treatment given above for the wearing ball will apply if $\lambda = \rho + 1$ is inserted.

Computational Procedures. The differential equations of dimensional percussive wear (for example, Eq 38 and 39) for the cases treated here, and for several others (such as two curved bodies, cylindrical impact), admit the "master curve" solution technique (Ref 1) (Fig. 12). A master curve (Fig. 13) is first constructed for a fixed initial condition, such as $\overline{N}_0 = 1$, $\overline{\lambda}_0 = 0.99$ (or $\overline{\rho}_0 = -0.01$, corresponding to $\lambda_0 - 1$).

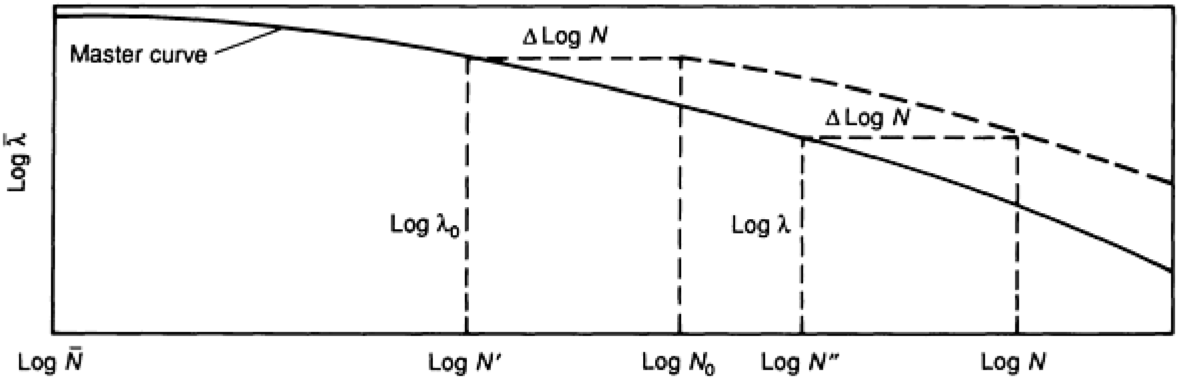


Fig. 12 Master curve shift procedure

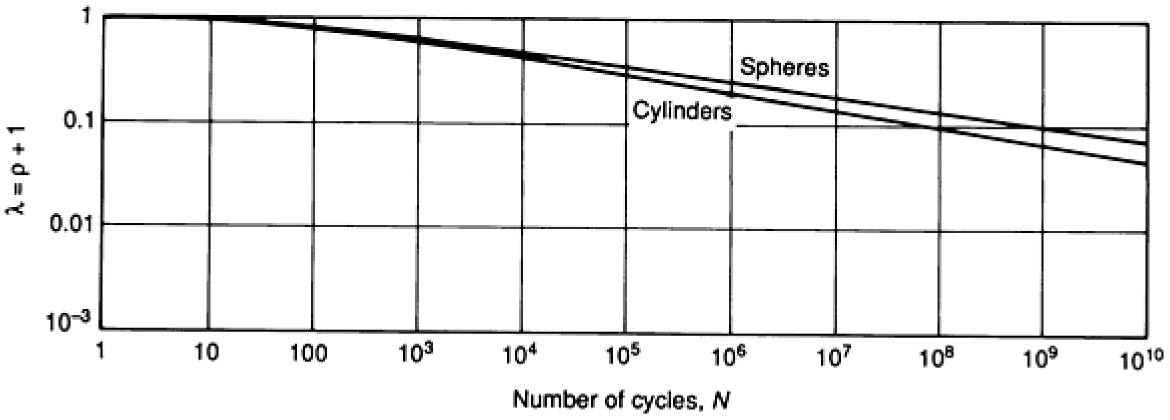


Fig. 13 Solution for impact wear in terms of nondimensional curvature parameters with initial conditions of N_0 = 1, ρ_0 = -0.01 or λ_0 = 0.99

In order to obtain the desired solution belonging to the given set of initial conditions (N_0, λ_0), we intercept the master curve at λ_0, obtaining the abscissa *N*'. Now, the master curve can be shifted from *N*' to N_0, and on this shifted portion of the master curve, the solution λ for any *N* desired can be found.

Besides the master curve method, an exact solution of Eq 39 can also be obtained for *N:*

$$N = \frac{N_0 \left(\frac{1-\lambda}{1-\lambda_0} \right)^{T-U}}{(\lambda/\lambda_0)^T} \quad \textbf{(Eq 43)}$$

where for a constant-impulse process, $T = (8 + 9c)/5$, $U = (3 + 9c)/5$.

Note that the compound nature of the impact (sliding component) is included in the zero-wear limit, and no further amendment of the computational apparatus is required. Several geometries that are more complex, and the case of two-body impact wear, treated in Ref 1. The calculation of hardness for a composite multilayered surface is treated in Ref 26.

Plotting a Wear Curve

By way of example, consider the wear of an originally plane, massive carbon steel machine platen, subjected to repetitive compound impact by a hard, nonwearing spherical-ended (*R* = 140 mm, or 5.5 in.) steel alloy (62 HRC) component of *m* = 1.2 g (0.04 oz) mass, at an approach speed of *V* = 2.0 m/s (78 in./s) and sliding speed of *v* = 1 m/s (40 in./s). The wearing surface has a roughness of $\delta/2 = 0.25\ \mu$m, and hardness of 20 HRC, with a yield stress, σ_y, of 500 MPa (75 ksi). The modulus of elasticity, *E,* of steel is 207 GPa (30×10^6 psi), and its Poisson's ratio, *v,* is 0.3. The friction coefficient, μ , for the unlubricated surfaces is 0.6.

Calculating the wear curve for nonlinear wear, the zero-wear limit is first established. The Hertz impact yields

$$P_0 = 929.4\text{N}$$

$$t^* = 9.494\ \mu\text{s}$$

$$K_1 = 245.9\ \text{MPa}$$

The slip factor is

$$f = \frac{\pi(1.2 \times 10^{-3})2}{(0.6)(929.4)(9.494 \times 10^{-6})} = 1.42$$

for a moderate range.

The β surface contribution factor is, from Fig. 5(c), $\beta = 320$, so that for the zero-wear limit, using $\sigma_y = H/3$ and $q = K_1$:

$$N_0 = \frac{1}{1 + 320}\left[\frac{(1.1)(500)}{245.9}\right]^9 2000 = 8743$$

and

$$h_0 = 0.25\ \mu\text{m}$$

Solving Eq 41 for *C*, 3.345 μm (130 μin.) is obtained, so that from Fig. 11, for $h_0/C = 0.0747$, the values obtained are $\lambda_0 = 0.93$ or $\rho_0 = -0.07$.

For the measurable wear range, the wear coefficient *c* = 3 will be used. Because the wear of a "soft plane" is being calculated, the nondimensional geometric wear parameter, $\lambda = \rho + 1$ will be computed. Equation 43 is now written:

$$N = 8743 \left(\frac{1 - \lambda}{1 - 0.93}\right) \ / \ \left(\frac{\lambda}{0.93}\right)^7$$

to be solved for $\lambda = 0.9, 0.8, 0.7, 0.6$, and so on, delineating the wear curve, Fig. 14. Conversion from nondimensional to dimensional wear depth, *h,* is made by use of the wear severity factor *C,* once again.

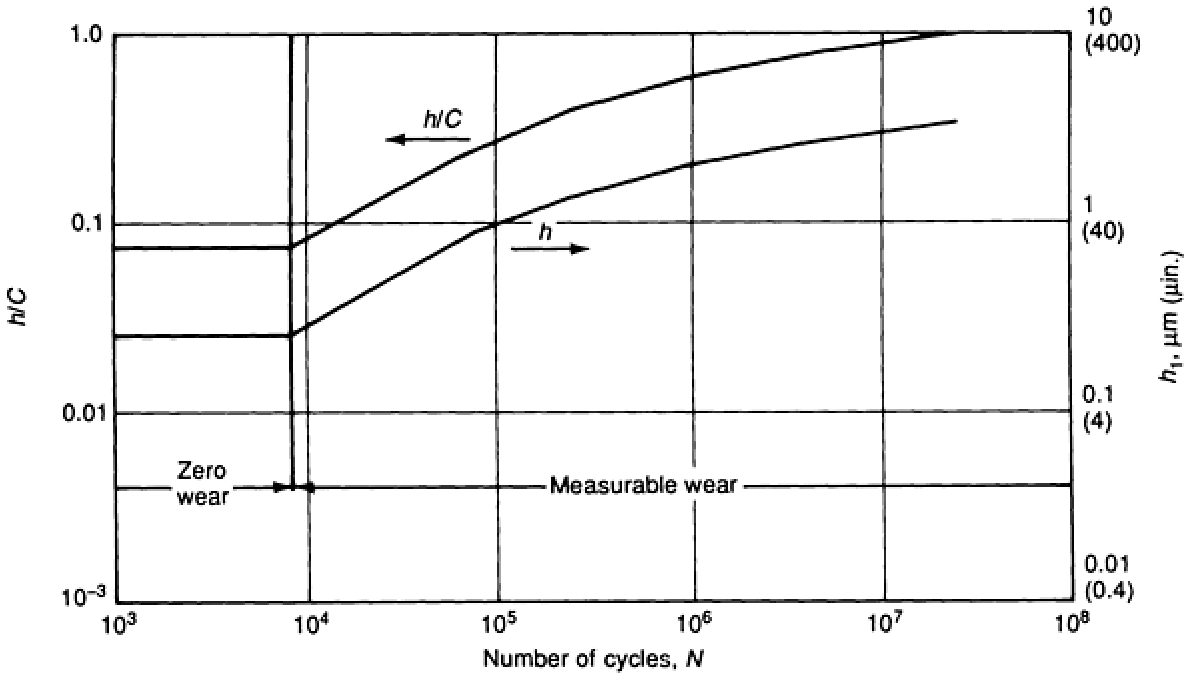


Fig. 14 Impact wear of platen by hard projectile; analytical prediction based on engineering data of example

References

1. P.A. Engel, *Impact Wear of Materials,* Elsevier, 1976
2. S.B. Ratner and E.E. Styller, Characteristics of Impact Friction and Wear of Polymeric Materials, *Wear,* Vol 73, 1981, p 213-234
3. P.A. Engel, H.C. Lee, and J.L. Zable, Dynamic Response of a Print Belt System, *IBM J. Dev.,* Vol 23 (No. 4), 1979, p 403-410
4. P.A. Engel *et al.,* Review of Wear Problems in the Computer Industry, *J. Lubr. Technol. (Trans. ASME),* Vol 24 (No. 100), 1978, p 189-198
5. A.W.J. DeGee, C.P.L. Commissaris, and J.H. Zaat, The Wear of Sintered Aluminum Powder (SAP) under Conditions of Vibrational Contact, *Wear,* Vol 7, 1964, p 535-559
6. P.L. Ko, The Significance of Shear and Normal Force Components on Tube Wear due to Fretting and Periodic Impacting, *Wear,* Vol 106, 1985, p 261-281
7. K. Wellinger and H. Breckel, Kenngroessen und Verschleiss beim Stoss Metallischer Werkstoffe, *Wear,* Vol 13, 1969, p 257-281
8. R.S. Montgomery, The Mechanism of Percussive Wear of Tungsten Carbide Composites, *Wear,* Vol 12, 1968, p 309-329
9. G.M. Sorokin, Types of Wear under Impact between Contacting Surfaces, *Mechanical Sciences* pamphlet UDC 539.538, (abridged English translation of *Mashinovedenie* by Scientific Information Consultants, Ltd.), Vol 3, 1974, p 49-52
10. R.G. Bayer, P.A. Engel, and J.L. Sirico, Impact Wear Testing Machine, *Wear,* Vol 19, 1972, p 343-354
11. P.A. Engel and J.L. Sirico, Impact Wear Study of Lubricated Contacts, *J. Lubr. (ASLE Trans.),* Vol 18, 1975, p 279-289
12. S.L. Rice, Reciprocating Impact Wear Testing Apparatus, *Wear,* Vol 45, 1977, p 85-95
13. S.L. Rice, The Role of Microstructure in the Impact Wear of Two Aluminum Alloys, *Wear,* Vol 54, 1979, p 291-301
14. S.L. Rice, H. Novotny, and S.F. Wayne, Characteristics of Metallic Subsurface Zones in Sliding and Impact

Wear, *Wear,* Vol 74, 1981-1982, p 131-142
15. T. Sugita, K. Suzuki, and Y. Nakata, Impact Wear of MgO Single Crystals, *Wear,* Vol 58, 1980, p 283-299
16. G. Levy and J. Morri, Impact Fretting Wear in CO Based Environments, *Wear,* Vol 106, 1985, p 97-138
17. I.R. Sare, Repeated Impact-Abrasion of Ore-Crushing Hammers, *Wear,* Vol 87, 1983, p 207-225
18. R.G. Bayer, Impact Wear of Elastomers, *Wear,* Vol 112, 1986, p 105-120
19. P.A. Engel and D.L. Millis, Surface Topography Study in Impact Wear, *Wear,* Vol 75, 1982, p 423-442
20. V. Veronesi, Wear Phenomena in Impact Printers--A Scanning Electron Microscopy Study, *Wear,* Vol 55, 1979, p 265-276
21. D. Tabor, *The Hardness of Metals,* Oxford University Press, 1951
22. P.A. Engel, J.L. Sirico, and T.H. Lyons, Impact Wear Theory for Steel Specimens, *Wear,* Vol 23, 1973, p 185-201
23. D.D. Roshon, Testing Machine for Evaluating Wear by Paper, *Wear,* Vol 30, 1974, p 93-103
24. T.C. Ku, R.G. Bayer, *et al., Handbook of Analytical Design for Wear,* Plenum, 1964
25. P.A. Engel, Predicting Impact Wear, *Mach. Des.,* Vol 49 (No. 12), 1977, p 100-105
26. P.A. Engel, A.R. Chitsaz, and E.Y. Hsue, Interpretation of Superficial Hardness for Multilayer, *Thin Solid Films,* to be published

Corrosive Wear

Brent W. Madsen, U.S. Department of the Interior, Bureau of Mines

Introduction

CORROSIVE WEAR is defined as the degradation of materials in which both corrosion and wear mechanisms are involved. The combined effects of wear and corrosion can result in total material losses that are much greater than the additive effects of each process taken alone, which indicates synergism between the two processes. This article focuses on the corrosion-wear synergism in aqueous slurry and grinding environments. Examples of corrosive wear problems encountered in industry are provided, and the effects of environmental factors on corrosive wear are discussed. Related articles can be found in this Volume. The article "Forms of Mechanically Assisted Degradation" in *Corrosion: Fundamentals, Testing, and Protection,* Volume 13A of *ASM Handbook,* is also relevant.

Occurrences in Practice

Corrosion and wear damage to materials, both directly and indirectly, costs the nation hundreds of billions of dollars annually. Although corrosion can often occur in the absence of mechanical wear, the converse is rarely true. Corrosion accompanies the wear process to some extent in all environments, except in a vacuum and inert atmospheres. Corrosion and wear often combine to cause aggressive damage in a number of industries, such as mining, mineral processing, chemical processing, and energy production. Corrosion and wear processes involve many mechanisms, the combined actions of which lead to the mutual reinforcement of their effectiveness. Seventeen synergistic relationships between abrasion, impact, and corrosion that could significantly increase wear damage in wet and aqueous environments have been identified (Ref 1).

Slurry Handling. The movement of materials as a slurry is an efficient means of transportation that is used in many industries. However, the movement of these slurries can cause significant corrosive wear. Pumps, elbows, tee junctions, valves, flotation cells, and hydrocyclones are parts of slurry handling systems that are subject to corrosive wear. The periodic replacement of worn parts often results in costly process downtime. Pipe rotation has been considered good practice for those pipes that are subject to localized wear, but careful monitoring is necessary to prevent failure of the worn parts because of corrosion (Ref 2). Turbulence and eddies are promoted either where fittings or valves project into the mainstream flow or when the diameter of a pipe changes. The momentum of the particles is sufficient to cause

abrasive damage to the side walls of the pipe in its curves or elbows, where the change in direction does not produce eddies. Slurry particles can impart damage to the bottom of the pipe by tumbling or sliding in straight pipe sections.

Sliding wear can occur, for example, in an ore chute in a minerals processing plant, where moist ore particles slide on an inclined platform, chute, or screen. Figure 1 shows a worn wire cloth screen in a minerals processing plant. Another example is a submerged valve that opens and closes to control the flow of material. Here, the parts are subjected to metal rubbing, as well as to the abrasive and corrosive nature of a slurry. Other, more common corrosive wear problems occur in equipment such as automobile brakes an engines (Ref 3), fillings used in dental applications, and tapes rubbed by recording heads in electronic video and audio equipment. Each of these wear examples has a corrosive component that results from exposure to reactive environments.

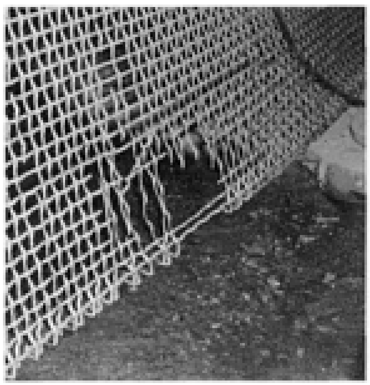


Fig. 1 Worn wire screen in minerals processing plant showing effects of sliding wear in moist environment

Crushing and Grinding. Corrosive wear is prevalent in the grinding of mineral ores. The annual cost of grinding media frequently approaches or exceeds the cost of energy in some processes (Ref 4). Figure 2 shows ball mill liner plates in both new and worn conditions. Figure 3 shows a cone crusher liner that wore completely through after only three months of use.

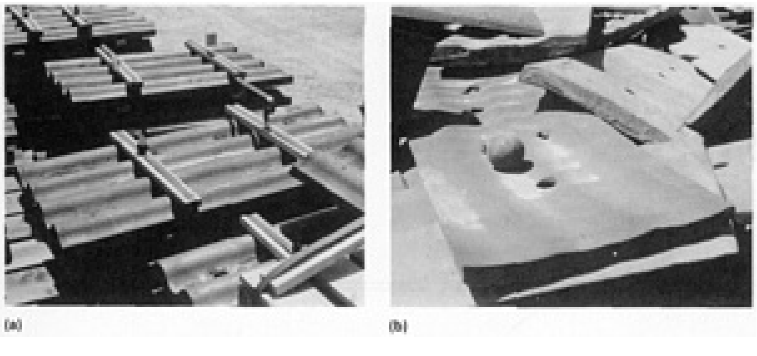


Fig. 2 Ball mill liner plates. (a) New condition, before installation. (b) Worn condition, with partially worn grinding belt

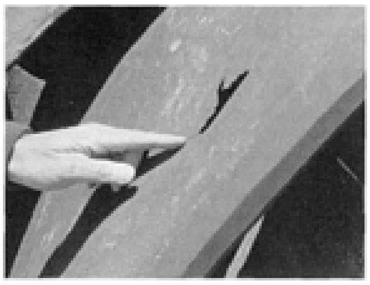


Fig. 3 Worn cone crusher liner after 3 months of use

Corrosion and wear damage arising from comminution is equivalent to 12% of more than 30 billion kW · h of energy consumed annually by U.S. industry for crushing and grinding purposes, whereas contributions from minerals account for 50%; cement, 25%; coal, 13%; and agricultural products, 12% (Ref 5). The development of more efficient mining, transportation, and processing systems has allowed exploitation of many large, low-grade ore bodies, which involves the crushing, grinding, and treatment of huge volumes of abrasive material to satisfy modern demand for materials (Ref 6). This trend will increase the future consumption of grinding media in an industry where over 0.9 Mg (1×10^9 tons) of ore are crushed annually (Ref 5).

Wear debris and corrosion products that are formed during comminution affect product quality and can adversely affect subsequent beneficiation by altering the chemical and electrochemical properties of the mineral system (Ref 4, 7, 8, 9). Electrochemical interactions between minerals and grinding media can occur, causing galvanic coupling that leads to increased corrosive wear. Corrosion and wear are also important in nonmining industries, such as the pulp and paper industry, where physical processing steps such as the grinding of wood chips are increasingly being substituted for chemical processing steps in order to reduce water-treatment requirements and fulfill regulatory standards.

High-Temperature Processes. Many chemical processes take place at elevated temperatures and involve corrosive wear. A process such as coal gasification involves hot gases with entrained solid particles that impinge on the containment vessel surface. Several studies have been conducted to describe the wear-corrosion synergism that takes place during the oxide formation and subsequent removal by abrasive particles (Ref 10, 11, 12).

Power-generation plants use processes that occur without any particles present and involve only the transport of solutions or steam. In these cases, liquid forces on the solid surface, which are due to turbulence or droplet impacts, mechanically remove protective layers of corrosion products, thus exposing more base material to corrosive action. An example of this type of process is the erosion-corrosion of steam turbines. Other systems may involve the dissolution of the protective oxide layer, because of a continual flow of liquid past the surface. Cavitation is also a problem where the high-speed flow of liquid is present. This subject is discussed in the article "Cavitation Erosion" in this Volume.

Effect of Environmental Factors on Corrosive Wear

Environmental factors affect corrosive wear in materials handling systems. Some of the more important factors that influence the material losses in slurry transport and grinding operations in aqueous media and ambient conditions are discussed below.

Slurry Particle Impingement on Two-Body Corrosive Wear

Slurry particles that strike the target material under the influence of their own momentum and that of the carrier fluid impart two-body wear damage to component parts. They are not constrained by another solid, as distinguished from grinding, which is discussed later.

Dependence on abrasive particle shape, density, size, and hardness, as related to wear rate, is discussed below.

Particle Shape. Until recently, no investigation had directly reported the effect of particle shape on wear in slurry handling systems. However, it was commonly agreed that the angularity of the slurry abrasives was important in determining the wear rate of materials where corrosion was not dominant. When slurries are recycled in a closed slurry

pot system, the wear rate can decrease dramatically with time (Ref 13, 14, 15). This is due to the microscopic rounding of the abrasive particles. The difference in wear rates for recycled and flow-through slurries is shown in Fig. 4 for type A514 steel in a water/silica sand slurry containing 50 × 70 mesh particles.

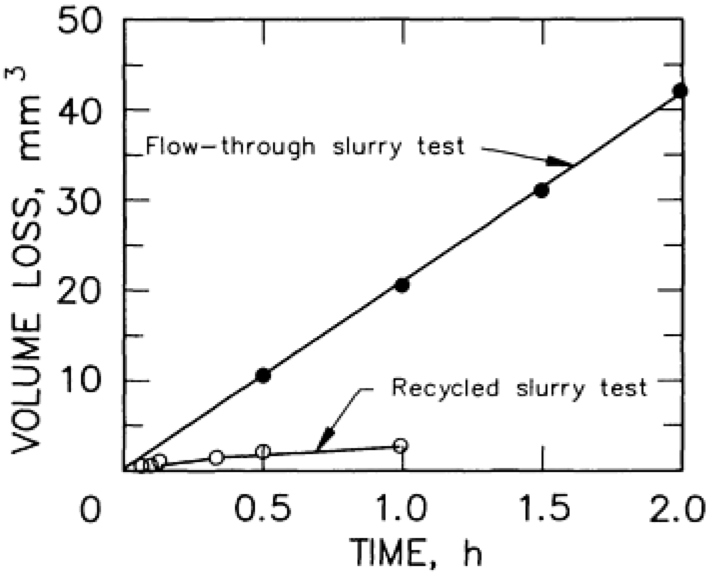


Fig. 4 Comparison of slurry wear for the flow-through and recycled slurry systems using a low-alloy steel in a 2% silica sand slurry. Source: Ref 13

Figure 5 shows the worn surfaces of the steel that resulted from the two tests. The specimen exposed for 1 h in the flow-through test (top photograph) had many fine grooves over its entire surface, typical of cutting wear. The bottom photograph shows the smooth, wave-type pattern that developed after 1.67 h with a recycled slurry, indicating a deformation wear mechanism.

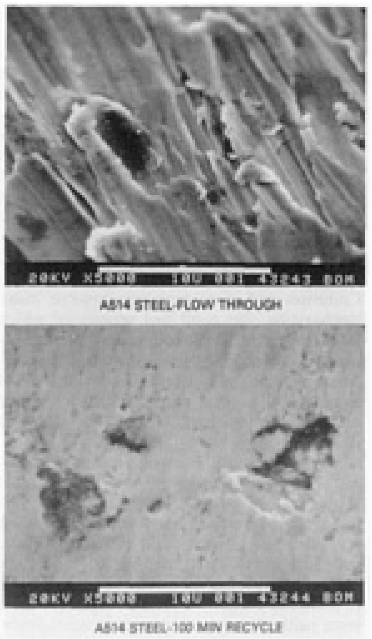


Fig. 5 Comparison of wear surfaces for low-alloy steel specimens worn in flow-through (top) and recycled (bottom) slurry tests for 1 h and 1.67 h, respectively. Source: Ref 13

The angularity of a particle depends on the radius of each protruding point on its surface, relative to its average radius. Bitter (Ref 16) has shown that the more angular a particle, the higher is its apparent density (Fig. 6). This effect results in a lower maximum velocity at which impact between the particles is purely elastic. If the predominant mode of material loss is corrosion, the angularity of the particles may not be such an important consideration.

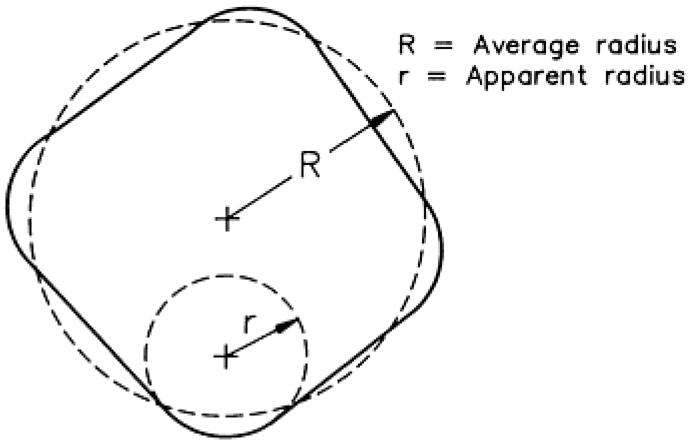


Fig. 6 Relationship between angularity and apparent density of a particle. Apparent density is $(R/r)^3$ × real reactive density. Source: Ref 16

Particle density is an important parameter that influences the corrosive wear caused by slurries. For a particle of a given angularity, the denser the particle, the more likely it is to cause either deformation or cutting wear on impact. This is a consequence of the dissipation of more energy in the same volume.

Particle Size. Because of their greater kinetic energy, particles of the same density and angularity in slurries with the same nominal velocity impart greater wear losses to materials as their size is increased. A recent study (Ref 17) has shown that larger particles have a higher collision efficiency and a higher impact velocity than smaller particles. These factors contribute to increased wear rates as slurry particle size increases.

Particle Hardness. The wear of materials has been shown to dramatically increase when the slurry particles are harder than the material being worn. However, increases in the hardness of the abrasive, above a critical hardness value, do not appreciably increase the wear rate. Likewise, when extremely hard abrasive particles are present in the slurry, moderate increases in the hardness of the worn material do little to retard the wear rate.

Dependence on slurry velocity, angle of attack, and solids concentration is discussed below in terms of wear rate.

Velocity of the slurry is the single most important factor that controls the rate of wear. The wear rate is an exponential function of velocity; exponents are reported to range from 1.6 to 4.8, but are generally between 2 and 3 (Ref 18). The velocity of the slurry not only affects the rate of mechanical damage of a material, but the corrosion rate, as well. Above a threshold velocity, corrosion products can effectively be stripped from an alloy, thus making available a new surface that is susceptible to corrosive attack.

The angle of attack of slurry particles on a material determines, to a large extent, the role that the abrasives will play in corrosive wear. In straight pipelines where the flow of slurry particles is basically parallel to the walls of the pipe, little damage to the underlying metal results from direct contact between the abrasive and alloy (Ref 19). The abrasives, along with the solution, serve to slowly remove the corrosion products and enhance the corrosion rate. The extent of this removal process is controlled by the particle size, angularity, and slurry density.

On the other hand, when a curvature in the pipe is present, the slurry particles tend to wear the pipe as depicted in Fig. 7. In this condition, the slurry particles are more likely to not only remove any corrosion products, but to directly impart mechanical wear damage to the alloy. If the alloy is one that normally passivates (such as a stainless steel), the abrasive

can continually remove the protective film and allow corrosion to proceed at a much higher rate than if no abrasive particles were present. The removal of the base metal depends on the type of material being worn.

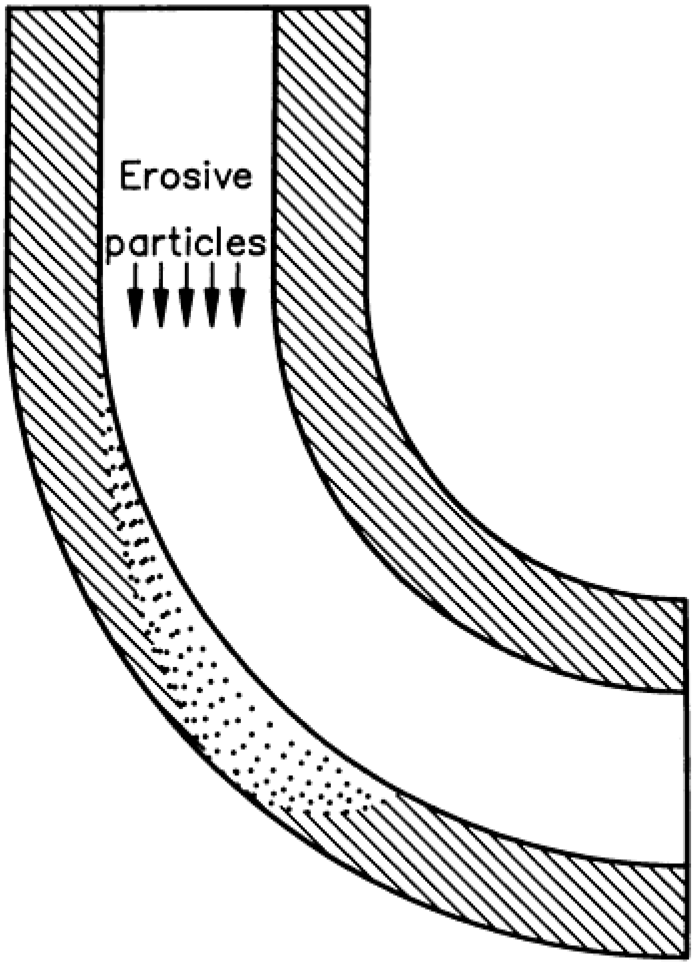


Fig. 7 Changing pattern of wear and, hence, changing angle of impingement, at bend in pipe. Source: Ref 16

Generally, ductile alloys wear greatest at shallow angles of attack, from 10 to 30°, whereas brittle materials wear greatest at 90°. The reason for these differences is the type of wear that occurs. Cutting wear is dominant for low angles of attack with ductile material, whereas deformation wear is greatest at larger angles of attack.

Solids Concentration. Increasing the solids concentration in a slurry generally increases the wear rate. The increase is only proportional to the solids concentration for dilute slurries. For denser slurries, particle-particle interaction tends to decrease the dependence of the wear rate on slurry density (Ref 14, 20).

Dependence on Hydrodynamics. Recent work (Ref 21, 22) on the trajectories and impact velocities of particles during slurry erosion has shown that not all of the particles directed at a target will impact it. The fraction of slurry particles that impact the surface of the target material is controlled by fluid velocity and viscosity, as well as the size, shape, and density of the slurry particles.

Flow-dependent corrosive wear typically occurs at geometrical irregularities, such as fittings, valves, and weld beads, where the flow separates from the wall of the containment vessel. Flow separation and reattachment produces high turbulence intensity and particle-wall interactions that lead to high corrosive wear rates (Ref 23, 24, 25). Figure 8 shows streamlines in a pipe for flow contraction and flow expansion.

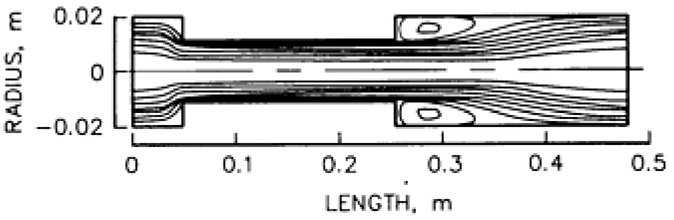


Fig. 8 Streamlines for water in pipe of variable diameter. Source: Ref 24

Blatt *et al.* (Ref 23) and Nesic *et al.* (Ref 24) have correlated material losses with the local, near-wall intensity of turbulence in single-phase (liquid only) flows. It was hypothesized that intensive near-wall turbulence disrupted the protective corrosion product and disturbed the mass transfer layer, thus enhancing oxygen transport to the metal surface and corrosion rates.

Corrosion products and the mass transfer of oxygen play an important role in the total material losses encountered in straight sections of a carbon steel pipeline. Work done by Postlethwaite *et al.* (Ref 19 and 26) shows that the material loss rate is under oxygen mass transfer control, with corrosion being the dominant mode of metal loss. The magnitude of the erosion-corrosion can be estimated using well-established mass transfer correlations for oxygen diffusion.

In aqueous solutions of near-neutral pH, mechanical wear prevents the formation of a rust film that completely covers the interior of a pipe. By providing a barrier to oxygen, this rust film in carbon steel pipes is responsible for corrosion rates of <1 mm/year (<0.04 in./year). When abrasive solids are present, coverage of the metal by the rust film is incomplete, and the islands of bare metal that are present can act as efficient cathodes and result in substantially increased rates, such as >10 mm/year (>0.4 in./year).

Grinding Wear: Impact and Three-Body Abrasive-Corrosive Wear

The wear of grinding media and crushers in mineral processing systems is caused by the combination of abrasion, corrosion, and high-energy impact of ore and metal components. The increased size of modern crushing and grinding equipment greatly increases the kinetic energy at metal/ore interfaces, which results in high wear rates. The crushed ore particles participate in three-body abrasion by virtue of being confined between two solid surfaces.

Dependence on Force. The strong effect of impact severity on the grinding media wear rate was reported by Dunn (Ref 1) for dynamic loading conditions. He showed that the wear rate was four times greater in an 8.5 m (28 ft) semiautogenous grinding (SAG) mill (8.30 kgf, or 18.3 lbf, balls) than that attained in a 2.9 m (9.5 ft) ball mill (1.81 kgf, or 4 lbf, balls). The ball velocities in the two mills were calculated to be 9.14 m/s (30 ft/s) and 5.33 m/s (17.5 ft/s), respectively. However, under conditions of a static load, Kotlyar *et al.* (Ref 27) found that the total wear rate of a metal alloy was approximately linear with load. They used a rotating metal alloy specimen between two ceramic anvils submerged in an abrasive slurry.

Dependence on Abrasive Type. Abrasive wear rates for steels and irons generally increase with an increase in mineral hardness. When the metal-to-mineral hardness ratio is greater than 0.6, marked improvement in the ability to resist abrasive wear is shown (Ref 28). In order to achieve this favorable hardness ratio, hard $(Fe,Cr)_3C$ carbides found in low-alloy nickel-chromium irons and harder $(Fe,Cr)_7C_3$ carbides found in high-chromium white cast irons are used.

Dependence on Galvanic Interaction between Minerals and Metal Alloys. Electrochemical interactions between ore materials and grinding media can occur. Mineral particles with potentials that are nobler than those of steel grinding media become galvanically coupled and result in accelerated corrosive wear (Ref 6). In recent studies (Ref 29, 30), it has been reported that the relative contributions from corrosive and abrasive wear are highly variable from one mineral-metal system to another, and are largely dependent on the mineral slurry characteristics, as well as the properties of the grinding media. In the presence of surface abrasion, and corrosion resistance, hardness of the grinding media, pH and conductivity of the pulp, and oxygen content influence the magnitude of galvanic currents between minerals and grinding media. Jang *et al.* (Ref 31) also found that galvanic currents exist between ferritic and martensitic phases in a microstructure of steel grinding media, as well as between sulfide minerals and these two phases.

Role of Localized Corrosion in Grinding. In the study of high-carbon low-alloy (HCLA) steel grinding media, Kotlyar and Wadsworth (Ref 27) found that localized electrochemical cells existed between strained and unstrained regions of the alloy. This resulted in preferential anodic dissolution in microgrooves, which promoted pitting. The overall corrosive wear rates of the steels were found to be strongly dependent on the localized corrosion.

Experimental Measurement of Corrosion-Wear Synergism

Quantitative measurements of corrosion, mechanical wear, and wear-corrosion synergism have been made using various experiments. The combined effects of wear, *W*, and corrosion, *C*, often result in total damage rates that are much greater than the additive effects of each process taken alone, indicating a strong synergism between wear and corrosion. The synergistic component, $S_{(C,W)}$ is that part of the total damage that results from the interaction of corrosion and wear processes. The total material loss, $T_{(C,W)}$, is related to the synergism by

$$T_{(C,W)} = W_0 + C_0 + S_{(C,W)} \qquad \textbf{(Eq 1)}$$

where W_0 is the wear damage in the absence of corrosion, and C_0 is the corrosion damage in the absence of mechanical wear. Equation 1 implies that corrosion can affect mechanical wear, and that mechanical wear can affect corrosion. Most researchers who have studied this type of system have measured *T* under conditions of corrosive wear, measured W_0 by conducting another experiment under conditions designed to minimize or eliminate corrosion during the wear process, and performed a third experiment to obtain the corrosion rate during the wear process. The rate obtained in this manner is usually higher than the corrosion rate C_0 because of mechanical wear interaction, and is designated as C_w. Using this notation, an equation similar to Eq 1 can be written as:

$$T_{(C,W)} = W_0 + C_w + S'_{(C,W)} \qquad \textbf{(Eq 2)}$$

where $S'_{(C,W)}$ is the increase in the mechanical wear losses when corrosion is present. Changes in the corrosion rate that are due to wear-corrosion synergism are included in C_w, which is the sum of all electrochemical corrosion processes occurring on the material during wear. The term W_0 can be obtained either by conducting the experiment under cathodic protection conditions or by using corrosion inhibitors.

The value C_w can be measured by performing polarization measurements with the specimen of interest as the working electrode. The electrochemical corrosion rate can be determined by making measurements of the current density within ±20 mV of the open-circuit corrosion potential. The current density versus polarization potential curves were fitted to the following theoretical equation that characterizes a reaction which contains only two processes under activation energy control:

$$I = I_c \left[\exp\left(2.203\frac{\phi}{\beta_a}\right) - \exp\left(-2.303\frac{\phi}{\beta_c}\right)\right] \qquad \textbf{(Eq 3)}$$

where *I* is the measured current density, I_c is the corrosion current density, ϕ is the polarization potential, and β_a and β_c are the anodic and cathodic Tafel slopes, respectively.

Equation 3 is suitable for describing the behavior of many alloys in aqueous solutions. A nonlinear regression technique can be used to determine I_c, β_a, and β_c, and values for the corrosion current can be converted to units of volume loss per area per hour to give C_w in Eq 2.

Another method used to obtain the corrosion rate is the polarization resistance method, in which the slope of the potential versus the current density (that is, the polarization resistance, R_p) is measured at potentials that are very close (±5 mV) to the open-circuit potential. The corrosion current, i_c, is calculated according to:

$$i_c = \frac{\beta_a \beta_c}{2.3(\beta_a + \beta_c)} \cdot \frac{1}{R_p} \quad \text{(Eq 4)}$$

The value I_c can then be calculated by dividing i_c by the electrode area. C_w (in units of material volume loss per area per unit time) is obtained after multiplying I_c by the appropriate conversion factors. In this case, β_a and β_c can either be measured graphically from Tafel plots of the potential versus log *i* or estimated from tabulated values for similar alloys.

Examples that show how these measurements have been made in slurry particle impingement systems and experiments that simulate ore-grinding environments are described below. Mechanisms for corrosive wear have been proposed, based on the cited studies.

Slurry Particle Impingement Systems

Various types of slurry wear tests designed to measure corrosive wear are reported in the literature and include slurry pot, jet impingement, and pipeline configurations. The experiments described in this section involve the wear condition known as low-stress, two-body abrasive wear. Madsen (Ref 13, 14, 32, 33) utilized a slurry pot that was capable of using either recycled or flow-through slurry. Pitt and Chang (Ref 34, 35) conducted experiments with a jet-impingement apparatus, and Postlethwaite *et al.* (Ref 26) used electrodes in a closed-loop pipeline to measure both wear and corrosion rates.

Slurry Pot Experiments. A slurry wear test developed by the Bureau of Mines (Ref 32, 33) was used to measure both wear and corrosion rates during slurry wear in order to establish the relationship between the wear and corrosion of metal-alloy specimens. Figure 9 depicts the slurry wear apparatus.

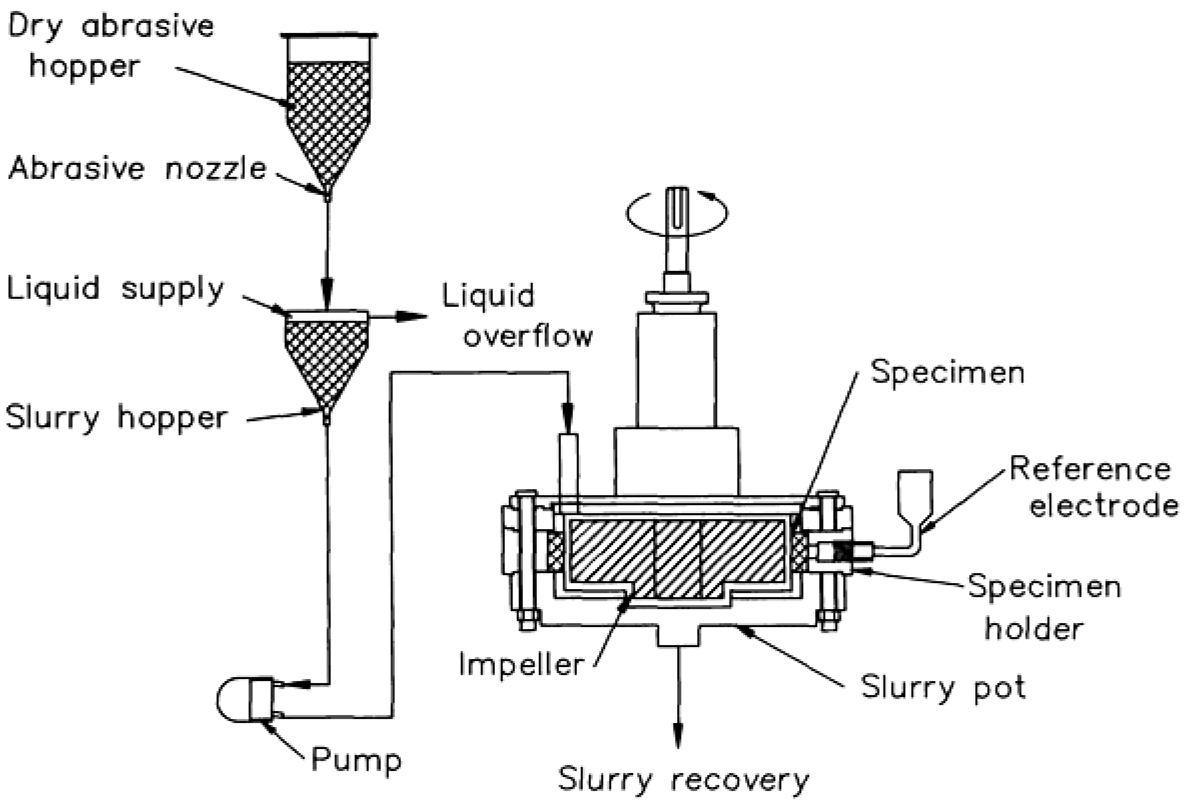


Fig. 9 Flow-through slurry wear apparatus. Source: Ref 13

Three types of tests were conducted to determine the terms in Eq 2. The total material loss, $T_{(C,W)}$, was determined from weight losses and densities during an experiment conducted at open circuit. The wear rate that was due to mechanical wear only was determined while cathodically protecting the specimens to eliminate corrosion.

These experiments showed that mechanical wear had a dramatic effect on the polarization curves of passive alloys, such as stainless steels. The results showed that the combined effects of wear and corrosion resulted in total wear rates that were much greater than the additive effects of each process taken alone. Figure 10 shows the polarization plots for 316 stainless steel that was exposed to a silica sand slurry at various slurry concentrations and impeller velocities. It is noted that the anodic current density increases with an increase in percent solids and speed, indicating that more of the passive film is removed because of the higher frequency of particle impacts. The oxide film on the stainless steel is apparently adherent and fast forming. In contrast, the polarization curves and corrosion rates of low-alloy steels were not affected to any extent because they freely corrode and do not form tightly adhering passive films.

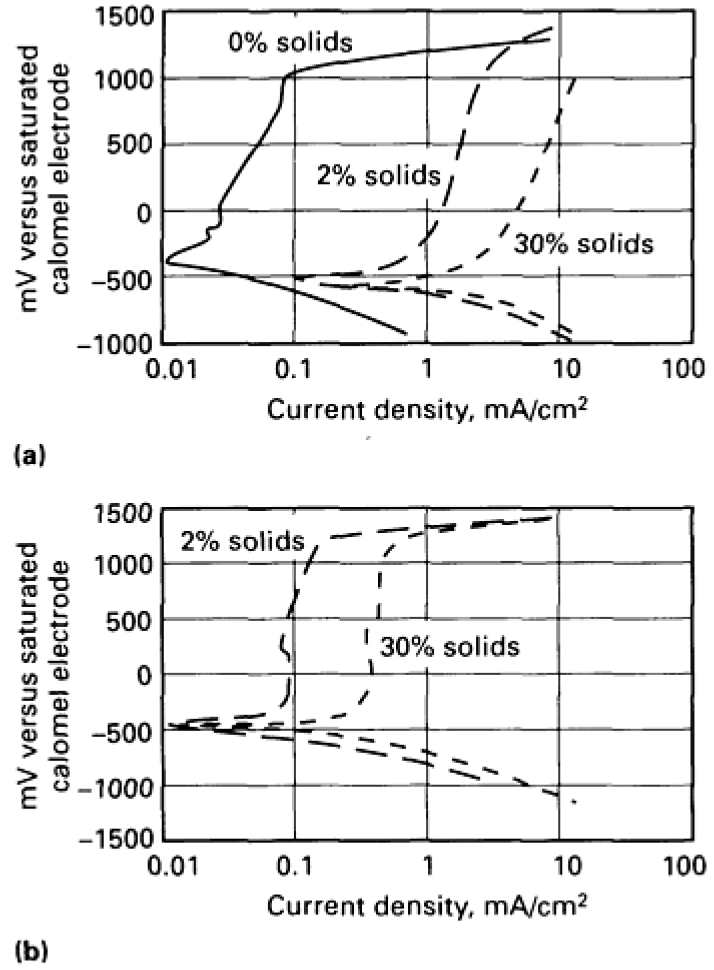


Fig. 10 Polarization curves for 316 stainless steel showing the effect of percent solids and impeller speed in a silica sand slurry. (a) Impeller velocity, 15.6 m/s (52 ft/s). (b) Impeller velocity, 5.8 m/s (19 ft/s). Source: Ref 33

The corrosive attack on the specimen surface can aid the mechanical wear processes, allowing the meal to be removed more easily by the forces that result from the abrasive action. In a 2% silica sand/water slurry, with impeller speeds of 5.8 and 15.6 m/s (19 and 51 ft/s), the $S'_{(C,W)}$ in Eq 2 accounted for 62% and 37%, respectively, of the total material losses, T. The synergistic effects are therefore very important in regard to the slurry wear of metal alloys.

Jet-impingement experiments were conducted in which Pitt and Chang (Ref 34, 35) used an electrochemically instrumented jet slurry apparatus. A hardened stainless steel nozzle with a 3.81 mm (0.15 in.) diam orifice delivered a slurry jet stream that impinged on the working electrode, which was 3 mm (0.12 in.) away from the orifice.

The corrosion currents were determined according to Eq 4 by using the polarization resistance technique described previously. $T_{(C,W)}$ was calculated from weight losses during corrosive wear. W_0 was determined from weight losses during experiments in which the working electrode was cathodically protected at a potential of -1.5 V versus the saturated

calomel electrode. This system used a recycled slurry that was replenished with fresh slurry periodically to minimize the effects that were due to slurry particle rounding.

At jetting speeds ranging from 9.3 to 21.5 m/s (30.5 to 70.5 ft/s), the $T_{(C,W)}$ from both a high-carbon steel and a high-chrome cast iron impinged by a quartz slurry were mainly influenced by the $S'_{(C,W)}$. This synergistic effect accounted for 8.3 to 54.9% of the total material losses, with the greatest effect occurring for larger particles and higher jet velocities.

Closed-Loop Pipeline Experiments. Postlethwaite *et al.* (Ref 26) measured the electrochemical corrosion rate of steels in a closed-loop pipeline experimental apparatus. He mounted electrode specimens flush with the pipe wall. The test section of pipe was a 38 mm (1.5 in.) diam vertical leg of the test loop. The vertical orientation of the pipe eliminated any effects of gravity on the solids concentration in the pipe.

The value for $T_{(C,W)}$ was obtained by weight losses, whereas C_w was obtained by polarization resistance techniques. The values obtained for W_0 (during cathodic protection of the electrode) were less than 5% of the total material losses, and $S'_{(C,W)}$ varied between 2.8 and 40.9% of the total material losses.

Most of the material losses from a commercial carbon steel pipe were due to electrochemical corrosion. Depending on the slurry flow rate, C_w varied from 56 to 92% of the $T_{(C,W)}$. It was shown that the erosive effect prevented the formation of a rust film that normally stifles the diffusion of oxygen to the corroding surface.

Grinding Wear Systems

The abrasive wear in grinding is different from that experienced in slurry particle impingement systems, because the abrasive particles from the crushed ore participate in three-body abrasion between two solid surfaces.

Rotating Cylinder/Anvil Experiments. Kotlyar, Pitt, and Wadsworth (Ref 36) used a rotating cylinder/anvil apparatus to measure the corrosive wear of a HCLA steel. The apparatus consisted of a rotating cylindrical specimen that rotated between two opposing abrasive anvils that were applied to the cylinder to a controlled load. The specimen served as a working electrode in a three-electrode system. Values of $T_{(C,W)}$ were determined by weighing the specimens before and after an experiment. To estimate W_0, cathodic protection was applied to the sample by fixing the voltage at -1.7 V versus a saturated calomel electrode (SCE). The electrochemical corrosion rate was determined using a linear polarization technique. In experiments using HCLA steel rotating cylinders in a 15 wt% quartz slurry (pH = 9), total material losses were higher by a factor of four when quartz, rather than HCLA, anvils were used.

As shown in Fig. 11, the increase in total material loss, T, was linear with applied load; corrosion, C_w, was nearly constant with load; the purely mechanical wear, W_0, increased and then leveled off as a function of load; and $S'_{(C,W)}$ was large and variable (Ref 37). The synergistic contribution of corrosion to abrasion increased markedly with loads higher than 65 N (6.6 kgf). This may result from the initiation and propagation of microcracks, assisted by anodic corrosion. The surface cracking caused by pitting and subsurface cracks (Ref 38) also has been observed after wet grinding in laboratory ball mills. The synergistic effect was found to decrease linearly with respect to pH in the range from 7 to 10. Above a pH of 10, pitting is retarded, and the metal loss that results from synergism decreases.

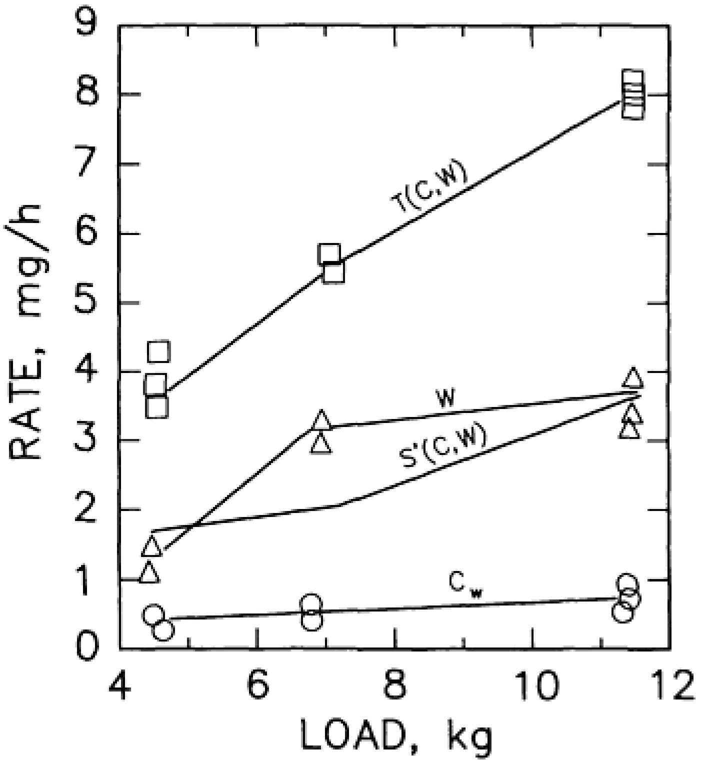


Fig. 11 Total wear rate, $T_{(C,W)}$; mechanical wear rate, *W;* corrosion rate during wear, C_w; and the synergistic term, $S'_{(C,W)}$, for HCLA steel using quartz anvils in a 15% quartz slurry at pH 9.0 as a function of load. Source: Ref 37

When localized corrosion is present, the corrosion component is small, whereas the synergistic component can be as high as 50% of the total material losses. The large synergism is accounted for by macroscopic removal of metal, resulting from localized corrosion.

Rotating Ball-on-Electrode Experiments. In investigations of the electrochemistry of complex sulfide-grinding systems, the galvanic currents of pyrite, pyrrhotite, and mild steel electrodes were determined in short-circuit connections under abrasive and nonabrasive conditions (Ref 39, 40). The device shown in Fig. 12was developed for the tests performed under abrasive conditions. It consisted of a porcelain ball rotating against three fixed electrodes of pyrite, pyrrhotite, and mild steel. Quartz slurries prepared with mild steel balls were used, and the galvanic current for each electrode was separately and alternately monitored via a switch with a zero-resistance ammeter.

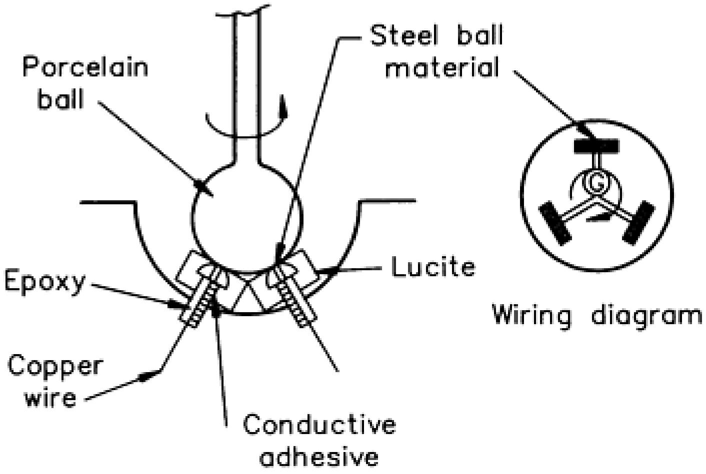


Fig. 12 Experimental ball-on-electrode device to measure galvanic currents between minerals and alloys. Source: Ref 38

Figure 13 depicts an electrochemical model for a two-sulfide/grinding media system according to the corrosion theory for a multielectrode galvanic cell (Ref 41). In this system, the noblest electrode is generally a sulfide mineral that is cathodic. The grinding media is usually more active than the sulfide minerals and is anodic. The other sulfide mineral develops an intermediate cathodic or anodic behavior, depending on its rest potential, the electrochemical characteristics of the electrodes, its surface area, and the material that it contacts.

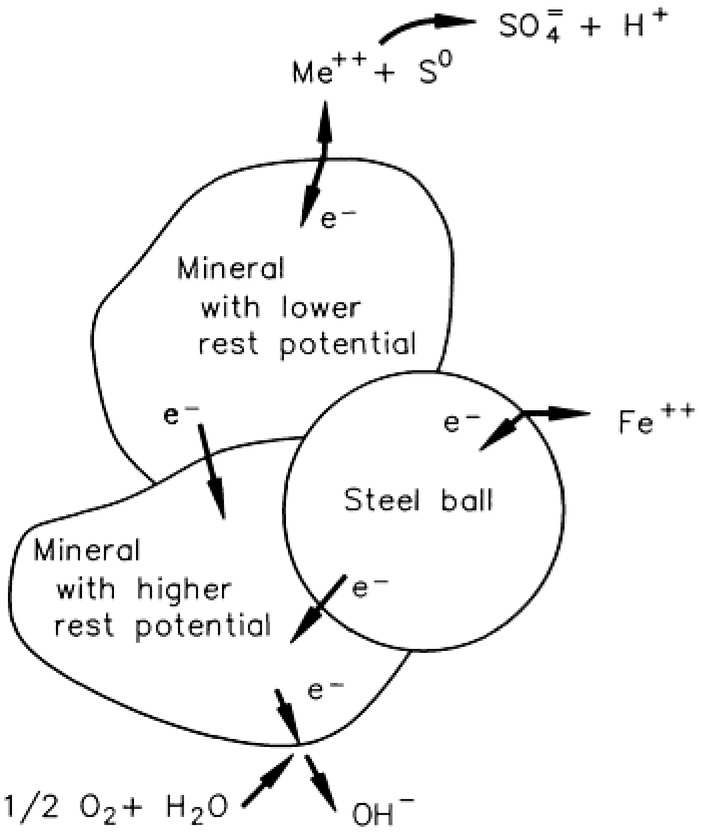


Fig. 13 Model of galvanic interactions among two minerals and grinding media. Source: Ref 39

Major reactions that are thought to take place at the surface of the respective mineral or steel electrode are (Ref 39):

$$\frac{1}{2}O_2 + H_2O + 2e^- = 2OH^- \text{ (at cathodic mineral)} \quad \textbf{(Eq 5)}$$

$$MeS = Me^{2+} + S + 2e^- \text{ (at anodic mineral)}; \quad S \rightarrow SO_4^{2-} + H^+ \quad \textbf{(Eq 6)}$$

$$Fe = Fe^{2+} + 2e^- \text{ (corrosion of steel media)} \quad \textbf{(Eq 7)}$$

$$Fe^{2+} + 2OH^- = Fe(OH)_2 \rightarrow \text{Fe hydroxides} \quad \textbf{(Eq 8)}$$

Dissociation of the minerals may also occur, but this is usually a slow process. The precipitation of other metal oxides is possible when dissolved metal ions from the minerals combine with hydroxide ions in a manner similar to Eq 8.

By using electrochemical and surface analysis techniques, Adam *et al.* (Ref 42) studied the pyrrhotite-grinding media system. They found a coating of hydroxide and/or sulfate species of iron on the pyrrhotite surface as a result of galvanic interactions (Eq 5, 7, 8) between the mineral and the grinding media. Similar findings were reported by Learmont and Iwasaki (Ref 43) for the galena-grinding media combination. The iron hydroxide coatings on pyrite produced by reactions (Eq 5, 7, 8) can impair floatability, whereas local pH changes that are due to acid formation via oxidation of elemental sulfur to sulfate and hydrogen ions at the pyrrhotite surface (Eq 6) can prevent the formation of the detrimental coating.

Corrosion currents that are due to the galvanic coupling between the sulfide minerals can be determined by obtaining polarization curves of individual electrodes under abrasive conditions, and by determining the point at which the cathodic curve crosses the anodic curve. These curves must be adjusted for the cathode/anode surface areas (Ref 44). Estimates based on this technique have shown good agreement with estimates from marked-ball water test data (Ref 9).

Mechanisms of Wear/Corrosion Synergism in Abrasive and Impact Wear

At stated earlier, Dunn (Ref 1) has indicated a number of mechanisms that account for wear/corrosion synergism in mineral processing systems. The following discussion applies the proposed mechanisms to various materials and handling systems.

Abrasion. The plastic deformation by high-stress metal-mineral contact causes strain hardening and susceptibility to chemical attack. This was shown by Kotlyar and Wadsworth (Ref 27), who demonstrated the existence of localized electrochemical cells between strained and unstrained regions of alloys, which promoted pitting. The abrasion mechanism itself can:

- Remove protective metal oxides and passive films to expose unoxidized metal to a corrosive environment. This is true for high-speed slurry wear of stainless steels (Ref 33). The corrosion rate of 316 stainless steel was shown to be as high as mild steel in the presence of only 2 wt% silica sand slurries
- Form microscopic grooves and dents that serve as sites for concentration cell corrosion. This can happen to many alloys, including HCLA steels, in which the formation of microgrooves and partial removal of oxide films lead to the initiation of pits that affect the total material losses (Ref 27)
- Increase the microscopic surface area exposed to corrosion, thus increasing the corrosion current
- Remove strain-hardened surface layers that are caused by repetitive impact. These layers are often more brittle than the underlying metal and do not adhere to the surface as well
- Crack brittle metal constituents, forming sites for impact hydraulic splitting. As corrosive liquids are trapped in cracks, forces on the metal surface seal the crack containing the liquid, and then hydraulically

split adjacent metal to propagate the crack.

Corrosion, as a mechanism, can:

- Produce pits that can be precursors to microcracking, which invites hydraulic splitting during impact
- Roughen the surface, thus leading to lower energy requirements to abrade the metal
- Produce hydrogen, which can embrittle the metal and lead to cracking (Ref 45)
- Selectively attack grain boundaries and less-noble phases of multiphase microstructure, weakening adjacent metal. This is especially true of white cast irons, where corrosion attacks the chromium-depleted zones that surround the hard carbides in the structure. This corrosion eventually promotes the premature removal of the harder carbide phases, thus accelerating material losses

Impact. Plastic deformation can make some constituents more susceptible to corrosion. The impact mechanism itself can:

- Crack brittle constituents and tear apart ductile constituents to form sites for crevice corrosion and hydraulic splitting
- Supply the kinetic energy to drive the abrasion mechanism, thus accelerating the abrasive component of wear. In ore processing, some minerals are driven into the softer steel grinding media, promoting the formation of galvanic corrosion cells
- Pressurize mill water to cause splitting, cavitation, and jet erosion of metal and protective oxidized material
- Pressurize mill water and gases to produce localized temperatures and phase changes in the liquid
- Heat grinding media, mill liners, ore, and fluids to increase corrosive effects

Means for Combating Corrosive Wear

Research and engineering efforts in wear, corrosion, and materials science are being pursued to combat corrosive wear in aqueous environments. The areas of investigation described below include materials selection, surface treatments, and handling environment modifications.

Materials Selection. The selection of the right material for a particular corrosive wear environment can lead to extended life of component parts, less costly downtime, and other economic advantages.

One approach is to place a number of materials in actual service and to compare the material losses of each over a given time. This technique can give the best solution to materials selection, but is time consuming. It also limits the number of materials that can be tried, and does not often result in the application of scientific principles to obtain the most cost-effective material for the given life of a process. However, actual field service should be used as a concluding step before final choices of materials are made.

Another approach is to adapt laboratory tests to field situations. Isaacson *et al.* (Ref 46) used a method in which electrodes were mounted at the ends of mill liner bolts, and polarization curves of the media material electrodes were obtained by a telemetry-radio system. Although the data were often noisy, correlations with laboratory corrosive wear tests were made.

Madsen (Ref 47, 48) used a test developed in the laboratory in a field experiment. He converted a laboratory slurry pot test to a portable model that could be transported to the field for use with slurries there. This approach enabled the testing of a large number of candidate materials in a relatively short time, because this slurry pot design allowed for the simultaneous comparison of 16 specimens.

The equipment could be useful in screening materials for further evaluation in actual service, thus shortening the time needed to determine the better material while offering a wider selection than actual field use could. This test is most

useful when the hydrodynamic and geometric parameters in the test chamber are similar to those experienced in the actual service conditions.

A more fundamental and scientific approach is to study the wear and corrosion characteristics of various materials in the laboratory and extrapolate these results to actual field use. A great deal of knowledge and modeling of wear, corrosion, 0and wear-corrosion synergism is needed, along with a detailed model of the hydrodynamics. This research is still in its infancy, but is making progress as the number of researchers who can recognize wear-corrosion synergism increases.

Surface Treatment. Rather than use the more-corrosive-resistant material to make the entire component, one may choose to use a thinner layer of a costly material on the surface of the more economical substrate material. Claddings, surface treatments such as hard facings and patching with welds, or replaceable liners have all been used. In pipelines where corrosion is a problem, it is sometimes economical to line the pipes with a smaller-diameter polymer that is more corrosion resistant, but lacks the strength to withstand high pressures on its own.

Modification of the materials handling environment can be effective in controlling the corrosive wear of component parts. Solution conditioning, such as adjusting the pH and deaeration, can reduce the amount of material losses in a corrosive-wear environment. Slurry conditioning is not economical for the short slurry lines used in mining operations unless some method of water separation and recirculation is used (Ref 19).

Use of Corrosion Inhibitors. Oxidizing inhibitors, such as chromates and nitrites, have been used to raise the potential of an alloy into passive regions and to lower their corrosion rates (Ref 49). These inhibitors form a surface film when used in high concentrations. Chromates also act as effective inhibitors when used at concentrations much lower than those required to produce active-passive transitions. At these lower concentrations, chromates act as cathodic inhibitors in neutral solutions. However, because chromates are toxic, alternative nontoxic inhibitors with a similar action should be considered. Many governments restrict the effluent limit to levels as low as 0.05 ppm chromate, which makes the choice of chromate inhibitors unacceptable. Progress in developing nonchromate inhibitors has been slow (Ref 19).

Slurry Parameters. Reduction of the slurry velocity is a major factor in controlling the rate of material losses if the mechanical wear, *W*, is important, because the wear rate generally varies with the velocity raised to exponents of 2, 3, or 4. For slurry pipelines that carry 20% silica sand (30 × 50 mesh), very little abrasive wear occurs below a velocity of 6 m/s (20 ft/s) (threshold value). Above this velocity, mechanical wear becomes more dominant as a means of material degradation.

Particle size can also be a factor in the wear of slurry handling equipment. The mechanical wear, *W*, will not be a problem if the particle size is sufficiently reduced so that the particles are fine enough to follow the streamlines of the solution, rather than impact the walls of the containment part. Postlethwaite points out, however, that fine slurry particles and low velocities may result in conditions mild enough to permit the growth of rust films and scale, which can lead to pitting (Ref 19).

The particle size and velocity combination should be maintained below the threshold value, where *W* is not a factor, but with a particle size that causes enough abrasion to maintain a rust-and scale-free pipe surface. In this condition, however, the pipe should be protected from corrosion by inhibitors and/or deaeration. This would eliminate the need for unnecessary size reduction.

Cathodic protection is a useful method for protecting short sections of slurry pipelines, pumps, elbows, and other equipment (Ref 50). However, the length of the protection must be kept sufficiently short to prevent overprotection and subsequent hydrogen blistering of the protected surface. The throwing power of cathodic protection, either by impressed current or galvanic anodes, is insufficient for this method to be used to protect the inside of slurry pipelines.

Sacrificial electrodes, such as zinc plugs in water or slurry pumps, have been used to cathodically protect the pump casting. This means of protection may be applicable to other instances where it is economically feasible.

Design of Pumps, Valves, Elbows. Research by Roco *et al.* (Ref 51, 52, 53, 54) and Ahmad *et al.* (Ref 55) has shown that computer modeling is becoming a tool for the design of wear-resistant slurry pumps. A computer code is used to carry out the numerical flow simulations within the pump channels, and an energy-based, wear-predictive model has allowed for the prediction of wear rates for various geometries.

The use of these types of models can extend the wear life of pumps by altering the geometry of the interior of the pump. This effectively removes the areas of high wear rates, and spreads the energy absorbed by the pump head evenly around its interior. This same type of modeling has been done by Postlethwaite (Ref 56) and Nesic and Postlethwaite (Ref 57, 58, 59) for areas of disturbed flow.

Other means of controlling corrosive wear are available, including:

- Increasing the pipe diameter in order to decrease the slurry velocity and help ensure laminar flow
- Increasing the thickness of materials in critical areas
- Inserting impingement plates or baffles to shield critical areas from high wear
- Directing the inlet flow of materials to avoid high particle velocities at the wall of containment vessels

References

1. D.J. Dunn, Metal Removal Mechanisms Comprising Wear in Mineral Processing, *Wear of Materials,* K.C. Ludema, Ed., American Society of Mechanical Engineers, 1985, p 501-508
2. A.M.F. Carter and D. Howarth, "A Literature Review of the Factors that Influence the Wear of Slurry-Handling Systems," Report M319, Council for Mineral Technology, Sept 1987
3. F.H. Stott, J.E. Breakell, and R.C. Newman, The Corrosive Wear of Cast Iron Under Potentiostatically-Controlled Conditions in Sulfuric Acid Solutions, *Corros. Sci.,* Vol 30, 1990, p 813-830
4. R.L. Pozzo and I. Iwasaki, Pyrite-Pyrrhotite Grinding Media Interactions and Their Effects on Media Wear and Flotation, *J. Electrochem. Soc.,* Vol 136 (No. 6), 1989, p 1734-1740
5. National Materials Advisory Board, "Comminution and Energy Consumption," NMAB-364, National Academy Press, 1981
6. K. Adam, K.A. Natarajan, S.C. Riemer, and I. Iwasaki, Electrochemical Aspects of Grinding Media-Mineral Interaction in Sulfide Ore Grinding, *Corrosion,* Vol 42 (No. 8), 1986, p 440-446
7. R.L. Pozzo and I. Iwasaki, Effect of Pyrite and Pyrrhotite on the Corrosive Wear of Grinding Media, *Miner. Metall. Process.,* Aug 1987, p 166-171
8. K.A. Natarajan, S.C. Riemer, and I. Iwasaki, Influence of Pyrrhotite on the Corrosive Wear of Grinding Balls in Magnetite Ore Grinding, *Int. J. Miner. Process.,* Vol 13, 1984, p 73-81
9. R.L. Pozzo and I. Iwasaki, An Electrochemical Study of Pyrrhotite-Grinding Media Interaction Under Abrasive Conditions, *Corrosion,* Vol 43 (No. 3), 1987, p 159-169
10. A.V. Levy and Y. Man, "Erosion-Corrosion Mechanisms and Rates in Fe-Cr Steels," Paper III, Corrosion 86 (Houston), National Association of Corrosion Engineers, 1986
11. S. Agarwal and M.A.H. Howes, Erosion/ Corrosion of Materials in High-Temperature Environments, *Proceedings of AIME Conference on High Temperature Corrosion in Energy Systems* (Detroit), American Institute of Mining, Metallurgical, and Petroleum Engineers, Sept 1984
12. T. Foley and A.V. Levy, The Erosion of Heat Treated Steels, *Wear,* Vol 48 (No. 1), 1983, p 181
13. B.W. Madsen and R. Blickensderfer, A New Flow-Through Slurry Erosion Wear Test, *Slurry Erosion: Uses, Applications, and Test Methods,* STP 946, J.E. Miller and F.E. Schmidt, Jr., Ed., ASTM, 1987, p 160-184
14. B.W. Madsen, A Study of Parameters Using a New Constant-Wear-Rate Slurry Test, *Wear of Materials 1985,* K.C. Ludema, Ed., American Society of Mechanical Engineers, 1985, p 345
15. R. Blickensderfer, J.H. Tylczak, and B.W. Madsen, "Laboratory Wear Testing Capabilities of the Bureau of Mines," IC 9001, Bureau of Mines, 1985
16. J.D.A. Bitter, A Study of Erosion Phenomena--Part II, *Wear,* Vol 6, 1963, p 169-190
17. R.S. Lynn, K.K. Wong, and H. Mcl. Clark, On the Particle Size Effect in Slurry Erosion, in *Wear of Materials 1991,* K.C. Ludema, Ed., American Society of Mechanical Engineers, 1991, p 77-82
18. H.Mcl. Clark, Slurry Erosion, *Proceedings of Conference on Corrosion-Erosion-Wear of Material at*

Elevated Temperatures (Berkeley), Electric Power Research Institute/ National Association of Corrosion Engineers, Jan 1990

19. J. Postlethwaite, The Control of Erosion-Corrosion in Slurry Pipelines, *Mater. Perform.,* Dec 1987, p 41-45
20. L.D.A. Jackson, Slurry Abrasion, *Trans. Can. Inst. Min. Metall.,* Vol 70, 1967, p 219-224
21. H. Hojo, K. Tsuda, and T. Yabu, Erosion Damage of Polymeric Material by Slurry, *Wear,* Vol 112, 1986, p 17-28
22. H. Hojo, K. Tsuda, and T. Yabu, Erosion Behavior of Plastics by Slurry Jet Method, *Kagaku Ronbunshu,* Vol 14, 1988, p 161-166
23. W. Blatt, T. Kohley, U. Lotz, and E. Heitz, The Influence of Hydrodynamics on Erosion-Corrosion in Two-Phase Liquid-Particle Flow, *Corrosion,* Vol 45 (No. 10), 1989, p 793-804
24. S. Nesic and J. Postlethwaite, Relationship Between the Structure of Disturbed Flow and Erosion-Corrosion, *Corrosion,* Vol 46 (No. 11), 1990, p 874-880
25. U. Lotz and J. Postlethwaite, Erosion-Corrosion in Disturbed Two Phase Liquid/ Particle Flow, *Corros. Sci.,* Vol 30 (No. 1), 1990, p 95-106
26. J. Postlethwaite, M.H. Dobbin, and K. Bergevin, The Role of Oxygen Mass Transfer in the Erosion-Corrosion of Slurry Pipelines, *Corrosion,* Vol 42 (No. 9), 1986, p 514-521
27. D. Kotlyar and M.E. Wadsworth, "The Role of Localized Corrosion on Corrosive Abrasive Wear of High Carbon Low Alloy Steel Grinding Media," Paper 246, Corrosion 88 (Houston), National Association of Corrosion Engineers, 1988
28. J.D. Watson, P.J. Mutton, and I.R. Sare, Abrasive Wear of White Cast Irons, *Met. Forum,* Vol 3 (No. 1), 1980
29. R. Perez and J.J. Moore, The Influence of Grinding Ball Composition and Wet Grinding Conditions on Metal Wear, *Wear of Materials 1983,* American Society of Mechanical Engineers, 1983, p 67-78
30. K.A. Natarajan and I. Iwasaki, Electrochemical Aspects of Grinding Media-Mineral Interactions in Magnetite Ore Grinding, *Int. J. Miner. Process.,* Vol 13, 1984, p 53-71
31. J.W. Jang, I. Iwasaki, and J.J. Moore, The Effect of Galvanic Interaction Between Martensite and Ferrite in Grinding Media Wear, *Corrosion,* Vol 45 (No. 5), 1989, p 402-407
32. B.W. Madsen, Measurement of Wear and Corrosion Rates Using a Novel Slurry Wear Test Apparatus, *Mater. Perform.,* Vol 26 (No. 1), 1987, p 21
33. B.W. Madsen, Measurement of Erosion-Corrosion Synergism with a Slurry Wear Test Apparatus, *Wear,* Vol 123, 1988, p 127-142
34. C.H. Pitt and Y.M. Chang, Electrochemical Determination of Erosive Wear of High Carbon Steel Grinding Balls, *Miner. Metall. Process.,* Aug 1985, p 166
35. Y.M. Chang and C.H. Pitt, Corrosive-Erosive Wear of Grinding Ball Metals at High Jet Velocities, *Corrosion,* Vol 43 (No. 10), 1987, p 599-605
36. D. Kotlyar, C.H. Pitt, and M.E. Wadsworth, Simultaneous Corrosion and Abrasion Measurements Under Grinding Conditions, *Corrosion,* Vol 44 (No. 5), 1988, p 221-228
37. C.H. Pitt, Y.M. Chang, M.E. Wadsworth, and D. Kotlyar, Laboratory Abrasion and Electrochemical Test Methods as a Means of Determining Mechanism and Rates of Corrosion and Wear in Ball Mills, *Int. J. Miner. Process.,* Vol 22, 1988, p 361-380
38. A.K. Gangopadhyay and J.J. Moore, "An Assessment of Wear Mechanisms in Grinding Media," presented at SME-AIME fall meeting, Society of Mining Engineers/ AIME, 1984
39. R.L. Pozzo and I. Iwasaki, Pyrite-Pyrrhotite Grinding Media Interactions and Their Effects on Media Wear and Flotation, *J. Electrochem. Soc.,* Vol 136 (No. 6), 1989, p 1734-1739
40. R.L. Pozzo, A.S. Malicsi, and I. Iwasaki, Pyrite-Pyrrhotite Grinding Media Contact and Its Effect on Flotation, *Miner. Metall. Process.,* Feb 1990, p 16-21
41. N.D. Tomashov, *Theory of Corrosion and Protection of Metals,* MacMillan, p 509-527
42. K. Adam, K.A. Natarajan, and I. Iwasaki, Grinding Media Wear and Its Effect on the Flotation of Sulfide Minerals, *Int. J. Miner. Process.,* Vol 12, 1984, p 39-54

43. M.E. Learmont and I. Iwasaki, Effect of Grinding Media on Galena Flotation, *Miner. Metall. Process.,* Aug 1984, p 136-144
44. I. Iwasaki, R.L. Pozzo, K.A. Natarajan, K. Adam, and J.N. Orlich, Nature of Corrosive and Abrasive Wear in Ball Mill Grinding, *Int. J. Miner. Process.,* Vol 22, 1988, p 345-360
45. E. Wandke and M. Moser, The Influence of Corrosion and Hydrogen Cracking on Blast Wear in Wet Media, *Wear,* Vol 121, 1988, p 15-26
46. A.E. Isaacson, P.J. McDonough, and J.H. Maysilles, "Corrosion Rate Determination in Industrial Ore Grinding Environments," Paper 232, Corrosion 88 (Houston), National Association of Corrosion Engineers, 1988
47. B.W. Madsen, A Portable Slurry Wear Test for the Field, *Trans. ASME--J. Fluids Eng.,* Vol 111, 1989, p 324-340
48. B.W. Madsen, A Comparison of the Wear of Polymers and Metal Alloys in Laboratory and Field Slurries, *Wear,* Vol 134, 1989, p 59-79
49. G.R. Hoey, W. Dingley, and C. Freeman, Corrosion Inhibitors Reduce Ball Wear in Grinding Sulfide Ore, *CIM Bull.,* Vol 68, 1975, p 120-123
50. J. Postlethwaite, B.J. Brady, M.W. Hawrylak, and E.B. Tinker, Effects of Corrosion on the Wear Patterns in Horizontal Slurry Pipelines, *Corrosion,* Vol 34 (No. 7), 1978, p 245
51. M.C. Roco, P. Nair, and G.R. Addie, Test Approach for Dense Slurry Erosion, *Slurry Erosion: Uses, Applications, and Test Methods,* STP 946, J.E. Miller and F.E. Schmidt, Jr., Ed., ASTM, 1987, p 185-210
52. M.C. Roco, "Wear Patterns in Centrifugal Slurry Pumps," Paper 224, Corrosion 88 (Houston), National Association of Corrosion Engineers, 1988
53. M.C. Roco, Optimum Wearing High Efficiency Design of Phosphate Slurry Pumps, *Proceedings of the 11th International Conference of Slurry Technology,* March 1986, p 277-285
54. M.C. Roco and P. Nair, Erosion of Concentrated Slurries in Turbulent Flow, *J. Pipelines,* Vol 4, 1984, p 213-221
55. K. Ahmad, R.C. Baker, and A. Goulas, Computation and Experimental Results of Wear in a Slurry Pump Impeller, *Proceedings of the Institution of Mechanical Engineers,* Vol 200 (No. C6), 1986
56. J. Postlethwaite, "Influence of Flow System Geometry on Erosion-Corrosion," Corrosion 91 (Houston), National Association of Corrosion Engineers, 1991
57. S. Nesic and J. Postlethwaite, Hydrodynamics of Disturbed Flow and Erosion-Corrosion, Part I--A Single-Phase Flow Study, *Can. J. Chem. Eng.,* Vol 69, 1991, p 698-703
58. S. Nesic and J. Postlethwaite, Hydrodynamics of Disturbed Flow and Erosion-Corrosion, Part II--A Two-Phase Flow Study, *Can. J. Chem. Eng.,* Vol 69, 1991, p 704
59. S. Nesic and J. Postlethwaite, A Predictive Model for Localized Erosion-Corrosion, *Corrosion,* Vol 47 (No. 8), 1991, p 582-589

Oxidational Wear

T.F.J. Quinn, School of Engineering and Applied Science, United States International University--Europe

Introduction

OXIDATIONAL WEAR is a form of wear that primarily occurs during nominally unlubricated conditions of sliding. It is usually, but not always, a beneficial form of corrosion in which oxide films that are approximately 3 or 4 μm (120 or 140 μin.) thick form on the real areas of contact at the relatively high "hot-spot" temperatures that often occur between sliding interfaces.

In many cases, tribological oxidation can reduce the wear rate by two orders of magnitude, compared with the wear of the same metals under an inert atmosphere. However, amelioration of the wear rate will only occur if the oxide layers are formed during sliding. It is not possible to artificially produce low-wear surfaces by preoxidizing the surfaces under static furnace conditions.

Oxidational wear can also occur under lubricated sliding conditions when the oil film thickness is less than the combined surface roughnesses of the triboelements, for example, under conditions of boundary lubrication. Oxidational wear is the "last defense" that a lubricated metal surface has against scuffing. It is also the only defense that some of the recently emerging ceramic tribosystems have against failure when running at high temperatures.

This article focuses on the dry, unlubricated wear of metals, particularly steels, because this reflects most of the work that has been conducted recently. Attention is also given to situations that could involve the mechanism of mild oxidational wear under conditions where oxide thickness increases parabolically with respect to time. The effect of the oxide thickness increasing linearly with respect to time is also discussed in view of the recent literature (Ref 1), which has shown that the wear of stainless steels may proceed in this way.

Theoretical expressions are given for the mild oxidational wear rate of a metal under conditions involving parabolic dependence of mass uptake of oxygen per unit area upon the time of oxidation, as well as conditions involving linear dependence. The parabolic oxidational wear rate is expressed in terms of a surface model that explicitly involves:

- The number of contacts, *N,* that constitute the real area of contact
- The equilibrium oxide film thickness, *TH,* beneath each of those contacts
- The temperature, *TF,* at the contacts
- The operating variables
- The parabolic oxidational constants of the metals being worn

On the other hand, the expression for the linear oxidational wear rate explicitly involves only the contact temperature, the operating variables, and the linear oxidation constants. This simplicity is only apparent, because the contact temperature is theoretically dependent on both the number of contacts and the thickness of the oxide beneath the contact.

It is difficult to confirm either of these expressions for the mild oxidation wear rate, because of the fact that the contact temperature is almost always impossible to measure directly. For most tribosystems, the contact temperature can only be deduced from direct measurement of the heat that flows away from the interface between each triboelement. If that is not possible, it can be deduced from theoretical analysis of the heat flow involving the number of contacts and oxide thickness.

Therefore, this article discusses the flow of heat from the sliding interface in terms of the same surface model that is used for the analysis of oxidational wear. Expressions are provided for the division of heat along the pin, *DTHP,* of a pin-on-disk wear configuration and for the excess temperature, *TE,* over the general surface temperatures of the pin, *TSP,* and of the disk, *TSD.* The *TE* is merely the difference between the *TF* and either the *TSP* or *TSD,* and is therefore not really independent of the expression for *DTHP.*

A selection of oxidational wear experiments is described, in which all of the tribologically important quantities (wear rates, frictional forces, divisions of heat, and general surface temperature) were determined experimentally using well-instrumented pin-on-disk wear machines with triboelements made of low-alloy steels. These steels are known to oxidize parabolically with time under both static furnace conditions and mild oxidational wear conditions (Ref 2).

The experimental wear rates and divisions of heat are then compared with the theoretical values predicted, respectively, by the parabolic oxidational wear equation and the equation for the theoretical division of heat along the pin. Because both equations depend on the same surface model of *N* asperity contacts (beneath which there is a critical oxide thickness, most of which has been formed at the contact temperature, which is itself dependent on *N* and oxide thickness), it should be possible to deduce values of the two unknowns, if one assumes that the static parabolic oxidation constants are relevant to tribological oxidation.

However, it has been found (Ref 3) that the only way to obtain a good correlation between the theoretical and experimental values of the wear rate and the division of heat is to assume that the tribological oxidation constants are different from the static oxidation constants. This article assumes that this difference resides entirely in the Arrhenius constant, *APT,* an assumption that might seem reasonable to make on purely physical grounds. However, there is a widely held belief, which is not necessarily correct, that the activation energy, *QP,* of oxidation in tribological contact is lower than that in the static furnace condition.

A computer program (FINDAP) is given in Table 1 for deducing the tribological Arrhenius constants from a selection of experiments using a low-alloy steel (En8). This program assumes that the static oxidation activation energies are the same as the tribological oxidation energies. Essentially, the program uses the method of "halving the interval" to obtain a value of *N,* which, when put into the equation for the division of heat, gives a value that is within 10^{-6} of the experimental value, *DE,* for a series of likely values of oxide thickness, *TH.*

Table 1 Details of FINDAP computer program

030 REM FINDAP1
050 REM PROGRAM PROVIDES VALUES OF TRIBOLOGICAL ARRHENIUS
060 REM CONSTANTS (AP) OBTAINED FROM ORIGINAL OXIDATIONAL WEAR
070 REM EQUATION IN WHICH VALUES OF NO. OF CONTACTS (N) AND
080 REM OXIDE THICKNESS (TH) ARE INSERTED. VALUES OF (N) ARE
100 REM OBTAINED BY COMPARING THEORETICAL AND EXPERIMENTAL
105 REM EXPRESSIONS FOR THE DIVISION OF HEAT (DTH) AND (DE).
110 REM THE THEORETICAL WEAR RATE ASSUMES THE
111 REM ACTIVATION ENERGY (QP) FOR OXIDATIONAL WEAR
112 REM EQUALS THE STATIC VALUE.
114 REM THE CHOICE OF (QP) DEPENDS UPON THE CONTACT
115 REM TEMPERATURE (TF), WHICH ITSELF DEPENDS
116 REM UPON (N) AND (TH).
117 REM
119 REM IT IS ASSUMED THAT (TH) MUST LIE WITHIN THE RANGE:
121 REM 0.5E-6m < TH < 5.0E-6m
122 REM AND (N) MUST LIE WITHIN THE RANGE:
123 REM 1 < N < 100000
125 DIM U(20),W(20),FF(20),TA(20),TB(20),TC(20),HT(20),XN(20)
130 DIM DE(20),TS(20),WR(20),H1(20)
140 LET XK=9.656E-6
141 REM XK IS THE THERMAL DIFFUSIVITY OF IRON
142 LET KO=2.1
143 REM KO=THE AVERAGE THERMAL CONDUCTIVITY OF IRON OXIDES
144 LET KS=37.2
145 REM KS=THE THERMAL CONDUCTIVITY OF STEEL
146 LET P=3.26E9
147 REM P IS THE ABSOLUTE HARDNESS OF THE STEEL (AT TS)
148 LET R=8.314
149 REM R IS THE UNIVERSAL GAS CONSTANT
150 LET KI=0.1045
151 REM KI IS THE THERMAL CONDUCTIVITY OF THE INSULATING
152 REM MATERIAL SURROUNDING THE PIN

```
153 LET RA=0.00795
154 REM RA IS THE INNER RADIUS OF THE COPPER CALORIMETER
155 LET RT=0.003175
156 REM RT IS THE RADIUS OF THE PIN
157 LET C=0.00001174
158 REM C IS THE HEAT FLOW RATE PER TEMPERATURE GRADIENT
159 REM ALONG THE THERMOCOUPLE WIRE MEASURING TA
160 LET L1=0.0067
161 REM L1 IS THE EXPOSED LENGTH OF PIN
162 LET L3=0.02
163 REM L3 IS THE INSULATED LENGTH OF THE PIN (BETWEEN
164 REM WHERE THE THERMOCOUPLES MEASURE TA AND TB)
170 LET M=SQR(2*KI/((KS*(RT^2))*LOG(RA/RT)))
171 REM M IS THE CONSTANT OF THE SECOND ORDER DIFFERENTIAL
172 REM EQUATION [T''] - [(M)^2][T] = 0
180 INPUT PROMPT "HOW MANY EXPERIMENTS TO BE ANALYZED": N
185 PRINT
190 FOR I = 1 TO N
192 PRINT
194 PRINT "PLEASE ENTER DETAILS OF EXPERIMENT NUMBER"; I
195 INPUT PROMPT "SPEED IN M/S": U(I)
196 PRINT "SPEED (IN M/S)= ";U(I)
197 INPUT PROMPT "LOAD IN N. ":W(I)
198 PRINT "LOAD (IN NEWTONS)= ";W(I)
199 INPUT PROMPT "FRICTION FORCE IN N. ";FF(I)
200 PRINT "FRICTION FORCE (IN NEWTONS)= ";FF(I)
201 INPUT PROMPT "TA IN DEGREES C ":TA(I)
202 PRINT "TA DEGREES C)= ";TA(I)
203 INPUT PROMPT "TB IN DEGREES C ":TB(I)
204 PRINT "TB (IN DEGREES C)= ";TB(I)
205 INPUT PROMPT "TC IN DEGREES C ";TC(I)
206 PRINT "TC (IN DEGREES C)= ";TC(I)
207 INPUT PROMPT "WEAR RATE (IN M^3/M) AS*.***E-** ":WR(I)
208 PRINT "WEAR RATE (IN M^3/M)= ";WR(I)
209 PRINT
210 NEXT I
220 FOR I = 1 TO N
222 REM (LINES 230 TO 282 RELATE TO DETERMINATION OF
223 REM (HEAT TRANSFER COEFFICIENT (H) IN TERMS OF
224 REM (REYNOLDS NUMBER (RE) & NUSSELDT'S NUMBER (RN)
230 LET RE=(2*RT*U(I)*1.165)/(1.9015E-05)
240 IF RE<=4 THEN 241
241 LET RN=0.891*(RE^0.33)
242 GOTO 282
250 IF RE<=40 THEN 251
251 LET RN=0.821*(RE^0.385)
252 GOTO 282
260 IF RE<=4000 THEN 261
261 LET RN=0.615*(RE^0.466)
262 GOTO 282
270 IF RE<=40000 THEN 271
271 LET RN=0.174*(RE^0.618)
272 GOTO 282
280 IF RE<=250000 THEN 281
281 LET RN=0.0239*(RE^0.805)
282 LET H=RN/0.2337
283 REM (LINES 291 TO 390 RELATE TO THE CALCULATOR OF
284 REM (THE EXPERIMENTAL DIVISION OF HEAT ALONG THE
285 REM (PIN (DE) AND THE GENERAL SURFACE TEMPERATURE
286 REM (OF THE PIN (TS). FOR DETAILS OF THE THEORY SEE
287 REM (PAPER BY QUINN,T.F.J., "REVIEW OF OXIDATIONAL
288 REM (WEAR - PART I: THE ORIGINS OF OXIDATIONAL WEAR"
289 REM (TRIBOLOGY INTERNATIONAL, VOL 16, p257, 1983)
291 LET Z=SQR(KS/(2*RT*H)
```

```
292 LET A=L1/(Z*RT)
300 LET B=TA(I)-TB(I)
301 LET D=TB(I)-TC(I)
302 LET E=TA(I)-TC(I)
310 LET CH=0.5*(EXP(M*L3)+EXP(-M*L3))
311 LET SH=0.5*(EXP(M*L3)-EXP(-M*L3))
320 REM CH=COSH and SH=SINH
330 LET H3=KS*PI*(RT^2)*M*((E*CH)-D)/SH
340 LET H2=((C*E)/(RA-RT)+H3
350 LET CH=0.5*(EXP(A)+EXP(-A))
351 LET SH=0.5*(EXP(A)-EXP(-A))
360 REM CH=COSH and SH=SINH
370 LET H1(I)=(PI*RT*(KS/Z)*(E*SH))+(H2*CH)
371 LET HT(I)=FF(I)*U(I)
380 LET DE(I)=H1(I)/HT(I)
390 LET TS(I)=(E*CH)+(Z*H2/(KS*PI*RT))*SH+TC(I)
400 NEXT I
401 FOR K = 1 TO N
404 PRINT
405 PRINT "ANALYSIS OF EXPERIMENT NO ";K
406 PRINT
407 PRINT
408 PRINT "SPEED(M/S)";U(K),"LOAD (N)";W(K)
409 PRINT "TA (C)";TA(K),"TB (C)";TB(K),"TC(C)";TC(K)
410 PRINT "WEAR RATE (M^3/M)";WR(K),"F.FORCE (N)";FF(K)
411 PRINT
412 PRINT "H1 (W)";H1(K), "HTOT (W)";HT(K), "DELTA EXPT";DE(K)
413 PRINT "BULK SURFACE TEMP TS (C)"; TS(K)
414 PRINT
415 REM
418 FOR TH=0.5E-6 TO 5.0E-6 STEP 0.5E-6
419 LET N1=1.0
420 LET N2=100000.0
422 LET N3=(N1+N2)/2.0
423 REM
424 REM
425 REM THIS IS THE METHOD OF HALVING THE INTERVAL
426 REM TO SOLVE THE EQUATION
427 REM DTH-DE(K)=0
428 REM FOR VARIOUS GIVEN VALUES OF OXIDE THICKNESS (TH)
429 REM
430 REM
431 DO
432 LET A1=SQR(W(K)/(PI*N1*P))
433 REM A1 IS THE RADIUS OF ASPERITY CONTACT FOR N=N1
434 LET L1=U(K)*A1/(2.0*XK)
435 REM L1 IS NOW DIFFERENT FROM DEFINITION IN LINE 160
436 LET TP1=HT(K)/(4.0*A1*KS*N1)+HT(K)*TH/(PI*A1*A1*KO*N1)
437 IF L1<=5.0 THEN 438
438 LET TD1=(-0.1021*L1+0.8605)*HT(K)/(4.0*A1*N1*KS)
439 IF L1>5.0 THEN 440
440 LET TD1=0.85*HT(K)/(4.0*A1*N1*KS*SQR(L1))
442 LET DTH1=TD1/(TD1+TP1)
443 LET FN1=(DTH1-DE(K))
444 REM TP1 IS [(TP)qrs] OF EQUATION (12)
445 REM TD1 (LINE 438) IS (TDL) OF EQUATION (15)
446 REM TD1 (LINE 440) IS (TDG) OF EQUATION (14)
447 REM LINES 449 TO 457 ARE REPEAT OF LINES 432
448 REM TO 443 WITH N=N2
449 LET A2=SQR(W(K)/(PI*N2*P))
450 LET L2=U(K)*A2/(2.0*XK)
451 LET TP2=HT(K)/(4.0*A2*KS*N2)+HT(K)*TH/(PI*A2*A2*(KO*N2)
452 IF L2<=5.0 THEN 453
453 LET TD2=(-0.1021*L2+0.8605)*HT(K)/4.0*A2*N2*KS)
```

```
454 IF L2>5.0 THEN 455
455 LET TD2=0.85*HT(K)/(4.0*A2*N2*KS*SQR(L2))
456 LET DTH2=TD2/(TD2+TP2)
457 LET FN2=(DTH2-DE(K))
458 REM LINES 460 TO 468 ARE A REPEAT OF LINES 432
459 REM TO 443 WITH N=N3
460 LET A3=SQR(W(K)/(PI*N3*P))
461 LET L3=U(K)*A3/(2.0*XK)
462 LET TP3=HT(K)/(4.0*A3*KS*N3)+HT(K)*TH/(PI*A3*A3*KO*N3)
463 IF L3<=5.0 THEN 464
464 LET TD3=(-0.1021*L3+0.8605)*HT(K)/4.0*A3*A3*N3*KS)
465 IF L3>5.0 THEN 466
466 LET TD3=0.85*HT(K)/(4.0*A3*N3*KS*SQR(L3))
467 LET DTH3=TD3/(TD3+TP3)
468 LET FN3=(DTH3-DE(K))
469 REM LINES 470 TO 479=APPLICATION OF THE HALVING ALGORITHM
470 IF (FN3*FN1)<0 THEN 471 ELSE 475
471 LET N2=N3
472 LET N1=N1
473 LET N3=(N1+N2)/2.0
474 GOTO 478
475 LET N1=N3
476 LET N2=N2
477 LET N3=(N1+N2)/2.0
478 LOOP UNTIL ABS(N1-N2)<0.000001 OR ABS(FN3)<=0.000001
479 LET N=N3
480 LET A=SQR(W(K)/(PI*N*P))
481 LET L=U(K)*A/(2.0*XK)
482 LET TP=HT(K)/4.0*A*KS*N)+HT(K)*TH(PI*A*A*KO*N)
483 IF L<=5.0 THEN 484
484 LET TD=(-0.1021*L+0.8605)*HT(K)/(4.0*A*A*N*KS)
485 IF L>5.0 THEN 486
486 LET TD=0.85*HT(K)/(4.0*A*N*KS*SQR(L))
487 LET DTH=TD/(TD+TP)
488 REM LET US NOW DEDUCE THE CONTACT TEMPERATURE (TF)
489 LET TF=DTH*TP+TS(K)
490 REM DIFFERENT QP VALUES FOR THE 3 POSSIBLE IRON OXIDES
492 IF TF<450.0 THEN 510
493 IF TF<600.0 THEN 530
494 REM CALCULATION OF AP FOR FEO
495 LET QP=210000.0
496 REM QP IS ACTIVATION ENERGY FOR FEO
497 LET F=0.2277
498 REM F IS THE FRACTION OF FEO THAT IS OXYGEN
499 LET RH=5.7E3
500 REM RH IS THE DENSITY OF FEO
501 LET APD=WR(K)*U(K)*P*TH*TH*F*F*RH*RH
502 LET APN=W(K)*(2.0)*A*EXP((-QP)/(R*(TF+273)))
504 LET AP=APD/APN
507 GOTO 540
508 REM LINES 510 TO 518 ARE A REPEAT
509 REM OF LINES 494 TO 502 FOR FE2O3
510 LET QP=208000.0
511 LET F=0.3006
512 LET RH=5.24E3
513 LET APD=WR(K)*U(K)*P*TH*TH*F*F*RH*RH
514 LET APN=W(K)*(2.0)*A*EXP((-QP)/(R*(TF+273)))
516 LET AP=APD/APN
518 GOTO 540
519 REM LINES 530 TO 536 ARE A REPEAT
520 REM OF LINES 494 TO 502 FOR FE304
530 LET QP=96000.0
532 LET F=0.2885
534 LET RH=5.21E3
```

536 LET APD=WR(K)*U(K)*P*TH*TH*F*F*RH*RH
537 LET APN=W(K)*(2.0)*A*EXP((-QP)/(R*(TF+273)))
538 LET AP=APD/APN
540 PRINT "OXIDE THICKNESS(M)";TH
542 PRINT "NUMBER OF CONTACTS (N)";N
544 PRINT "OXIDATION TEMPERATURE(TO)";TF
546 PRINT "THEORETICAL DIVISION OF HEAT";DTH
548 PRINT "ACTIVATION ENERGY";QP
549 PRINT "ARRHENIUS CONSTANT";AP
550 PRINT "VIRTUAL PIN TEMP (TP)";TP
551 PRINT "VIRTUAL DISC TEMP (TD)";TD
552 PRINT "RADIUS OF CONTACT (a)";A
553 PRINT
554 PRINT
555 NEXT TH
556 NEXT K
560 STOP
570 END

These pairs of values of *N* and *TH* are then inserted into the mild oxidational wear equation to eventually provide a value of *APT* that, together with the value for *QP* that is appropriate for the calculated value of *TF,* is consistent with the experimentally measured wear rate.

With the tribological Arrhenius constants obtained from FINDAP, another numerical method is used as the basis of a computer program (OXYWEAR, Table 2) to deduce the critical oxide thickness, number of contacts, and contact temperature for a set of wear experiments with a different low-alloy steel. The basis of the numerical method is to: express *N, TH,* and *TF* in terms of the contact radius, *A;* obtain a quadratic equation in *A;* and thereby solve for *A* using a fully iterative technique. It turns out that the values obtained for *N, TH,* and *TF* are very similar to those arising from the FINDAP program, thereby indicating the credibility of the method.

Table 2 Details of OXYWEAR computer program

100 REM OXYWEAR PROGRAM TO SOLVE OXIDATIONAL WEAR EQUATION
101 REM FOR NORMAL HOT-SPOT CONDITIONS BY A FULLY ITERATIVE
102 REM TECHNIQUE
105 DIM U(20),W(20),FF(20),TA(20),TB(20),TC(20),HT(20)
106 DIM DE(20),TS(20),WR(20),H1(20)
110 LET XK-9.656E-6
112 REM XK IS THE THERMAL DIFFUSIVITY OF THE DISC MATERIAL
114 LET KO=2.1
115 REM KO IS THE THERMAL CONDUCTIVITY OF THE OXIDE
118 LET KS=37.2
120 REM KS IS THE THERMAL CONDUCTIVITY OF THE STEEL
122 LET P=3.26E9
124 REM P=BULK HARDNESS OF THE STEEL AT TEMPERATURE (TS)
125 REM (IN UNITS OF STRESS)
126 LET R=8.314
128 REM R IS THE UNIVERSAL GAS CONSTANT
130 LET KI=0.1045
132 REM KI=THERMAL CONDUCTIVITY OF THE INSULATING MATERIAL
134 LET RA=0.00795
136 REM RA IS THE INNER RADIUS OF THE COPPER CALORIMETER
138 LET RT=0.003175
140 REM RT IS THE RADIUS OF THE PIN
142 LET C=0.00001174
144 REM C = HEAT FLOW RATE PER TEMPERATURE GRADIENT ALONG
145 REM THE THERMOCOUPLE MEASURING (TA)
146 LET L1=0.0067
148 REM L1 IS THE EXPOSED LENGTH OF PIN
150 LET L3=0.02
152 REM L3 IS THE INSULATED LENGTH OF PIN

```
154 LET M=SQR(2*KI/((KS*(RT^2))*LOG(RA/RT)))
156 REM M=CONSTANT OF THE SECOND ORDER DIFFERENTIAL EQUATION
157 REM GOVERNING THE HEAT FLOW i.e. T'' - [(M)^2][T] = 0
160 REM
180 INPUT PROMPT "HOW MANY EXPERIMENTS TO BE ANALYSED ": N
185 PRINT
190 FOR I = 1 TO N
192 PRINT
194 PRINT "PLEASE ENTER DETAILS OF EXPERIMENT NUMBER";I
195 INPUT PROMPT "SPEED IN M/S ":U(I)
196 PRINT "SPEED (IN M/S)=":U(I)
197 INPUT PROMPT "LOAD IN N. ":W(I)
198 PRINT "LOAD (IN NEWTONS)=";W(I)
199 INPUT PROMPT "FRICTION FORCE IN N. ":FF(I)
200 PRINT "FRICTION FORCE (IN NEWTONS)=";FF(I)
201 INPUT PROMPT "TA IN DEGREES C":TA(I)
202 PRINT "TA (IN DEGREES C) =";TA(I)
203 INPUT PROMPT "TB IN DEGREES C ":TB(I)
204 PRINT "TB (IN DEGREES C) =";TB(I)
205 INPUT PROMPT "TC IN DEGREES C ":TC(I)
206 PRINT "TC (IN DEGREES C)= ";TC(I)
207 INPUT PROMPT "WEAR RATE (IN M^3/M) AS*.***E-**":WR(I)
208 PRINT "WEAR RATE (IN M^3/M)= ":WR(I)
209 PRINT
210 NEXT I
220 FOR I = 1 TO N
221 REM
222 REM LINES 230 TO 290 RELATE TO THE CALCULATION OF HEAT
223 REM TRANSFER COEFFICIENT (H) IN TERMS OF REYNOLDS NUMBER
224 REM (RE) AND NUSSELDT'S NUMBER (RN)
225 REM
230 LET RE=(2*RT*U(I)*1.165)/(1.9015E-05)
240 IF RE<=4 THEN 241
241 LET RN=0.891*(RE^0.33)
242 GOTO 290
250 IF RE<=40 THEN 251
251 LET RN=0.821*(RE^0.385)
252 GOTO 290
260 IF RE<=4000 THEN 261
261 LET RN=0.615*(RE^0.466)
262 GOTO 290
270 IF RE<=40000 THEN 271
271 LET RN=0.174*(RE^0.618)
272 GOTO 290
280 IF RE<=250000 THEN 281
281 LET RN=0.0239*(RE^0.805)
290 LET H=RN/0.2337
291 LET Z=SQR(KS/(2*RT*H))
292 LET A=L1/(Z*RT)
300 LET B=TA(I)-TB(I)
301 LET D=TB(I)-TC(I)
302 LET E=TA(I)-TC(I)
310 LET CH=0.5*(EXP(M*L3)+EXP(-M*L3))
311 LET SH=0.5*(EXP(M*L3)-EXP(-M*L3))
320 REM CH=COSH and SH=SINH
330 LET H3=KS*PI*(RT^2)*M*((E*CH)-D)/SH
331 REM H3 IS EQUATION (11) OF REVIEW PAPER (17)
340 LET H2=((C*E)/(RA-RT))+H3
341 REM H2 IS EQUATION (12) OF REVIEW PAPER (17)
350 LET CH=0.5*(EXP(A)+EXP(-A))
351 LET SH=0.5*(EXP(A)-EXP(-A))
360 REM CH=COSH and SH-SINH
365 LET H1(I)=(PI*RT*(KS/Z)*(E*SH))+(H2*CH)
366 REM H1 IS EQUATION (14) OF REVIEW PAPER (17)
```

```
371 LET HT(I)=FF(I)*U(I)
372 REM HT IS THE TOTAL HEAT EVOLVED AT THE CONTACT
380 LET DE(I)=H1(I)/HT(I)
381 REM DE IS EQUATION (15) OF REVIEW PAPER (17)
390 LET TS(I)=(E*CH)+((Z*H2/(KS*PI*RT))*SH)+TC(I)
391 REM TS IS THE GENERAL SURFACE TEMPERATURE OF THE PIN
392 NEXT I
393 FOR K = 1 TO N
394 LET H5=INT(1(K)*1E4+0.5)/1E4
395 LET H6=INT(HT(K)*1E4+0.5)/1E4
396 LET D1=INT(DE(K)*1E4+0.5)/1E4
397 LET T1=INT(TS(K)*1E4+0.5)/1E4
398 REM LINES 394 THROUGH 397 ENSURE THAT H1,HT,DE and TS
399 REM ARE GIVEN TO FOUR DECIMAL PLACES
400 LET RD=2.0E-06
401 REM RD IS FIRST TRIAL VALUE OF THE CONTACT RADIUS (R1)
402 LET APT=6.12E12
403 REM APT IS THE TRIBOLOGICAL ARRHENIUS CONSTANT FOR THE
404 REM FORMATION OF FE2O3. IT HAS BEEN OBTAINED BY APPLYING
405 REM THE FINDAP COMPUTER PROGRAM TO SOME WEAR RESULTS
406 REM WITH EN8 STEEL
407 LET QP=208000
408 REM QP=THE OXIDATIONAL ACTIVATION ENERGY FOR THE STATIC AND
409 REM TRIBOLOGICAL OXIDATION FOR FE2O3
410 LET F=0.3006
411 REM F IS THE FRACTION OF FE2O3 THAT IS OXYGEN
412 LET RH=5.24E03
413 REM RH IS THE DENSITY OF FE2O3
414 LET I=-1
415 PRINT
416 PRINT "ANALYSIS OF EXPERIMENT NO ";K
417 PRINT
418 PRINT
419 PRINT "SPEED(M/S) "; U(K), "LOAD (N) ";W(K)
420 PRINT "TA (C)";TA(K),"TB (C)";TB(K),"TC(C)";TC(K)
422 PRINT "WEAR RATE (M 3/M)" ;WR(K),"F.FORCE (N)";FF(K)
423 PRINT
424 PRINT "H1 (W)" ; "HTOT (W)";H6, "DELTA" ;D1
425 PRINT "BULK SURFACE TEMP TS (C)" ;TS(K)
426 PRINT
427 PRINT "IF FE2O3 THEN:-"
430 LET B=(PI*KO)/(4*DE(K)*KS)
431 REM B IS GIVEN BY EQUATION 20(b) OF REVIEW PAPER (17)
432 LET C=(0.8605*(1-DE(K))-DE(K))
433 REM C IS GIVEN BY EQUATION 20(c) OF REVIEW PAPER (17)
434 LET E=(0.1021*U(K)*(1-DE(K)))/(2*XK)
436 REM E IS GIVEN BY EQUATION 20(d) OF REVIEW PAPER (17)
438 LET G=DE(K)*HT(K)*P*((PI/(4*W/(K)*KS))+((B*C)/(W(K)*KO)))
440 REM G IS GIVEN BY EQUATION 21(b) OF REVIEW PAPER (17)
450 LET RI=(DE(K)*HT(K)*P*B*E)/(KO*W(K))
452 REM RI IS GIVEN BY I IN EQ. 21(c) OF REVIEW PAPER (17)
460 LET RJ=(2*W(K)*APT)/(U(K)*F*F*RH*RH*B*B)
462 REM RJ IS GIVEN BY J IN EQ 22(b) OF REVIEW PAPER (17)
470 LET RM=R*G
472 REM RM IS GIVEN BY M IN EQ 22(c) OF REVIEW PAPER (17)
474 LET S=R*RI
476 REM S IS GIVEN BY EQUATION 22(d) OF REVIEW PAPER (17)
482 LET V=R*(TS(K)+273)
483 REM V IS GIVEN BY EQUATION 22(e) OF REVIEW PAPER (17)
484 LET Z1A=C*C*WR(K)*RD
485 LET Z1B=2*C*E*WR(K)*RD*RD
486 LET Z1C=E*E*WR(K)*RD*RD*RD
487 LET Z1=Z1A-Z1B+Z1C
488 REM Z1/WR(K) IS THE DIVIDEND OF THE PRE-EXPONENTIAL
```

489 REM TERM IN THE EQUATION FOR THE WEAR RATE (i.e.EQ.22(a)
490 REM OF REVIEW PAPER (18))
500 ON SGN(RJ)+2 GOTO 660,660,510
510 ON SGN(Z1)+2 GOTO 660,660,520
520 LET Z2=LOG(Z1)-LOG(RJ)
530 LET R1=-(QP/(Z2*RM))-(V/RM)+(S*RD*RD/RM)
535 REM R1 IS THE (R+1)th VALUE OF THE CONTACT RADIUS (A)
536 REM WHERE RD IS THE Rth VALUE (FOR STARTING VALUE - SEE
537 REM 400) AND IS GIVEN BY EQUATION (A2) OF PAPER BY
538 REM QUINN, ROWSON AND SULLIVAN (REF 2)
540 IF ABS(R1-RD)<=(RD/1000) THEN GOTO 560
550 LET RD=R1
551 GOTO 486
560 LET N=W(K)/(PI*P*R1*R1)
561 LET TH=B*R1*(C-(E*R1))
562 REM TH IS THE OXIDE THICKNESS AS GIVEN BY EQUATION 20(a)
564 REM OF REVIEW PAPER (17)
568 LET TF=(G*R1)-(RI*R1*R1*)+TS(K)
569 REM TF IS THE CONTACT TEMPERATURE AS GIVEN BY EQUATION
570 REM 21(a) OF REVIEW PAPER (17)
572 LET N=INT(N*10+0.5)/10
573 LET TF=INT(TF*10+0.5)/10
580 PRINT "OXIDE THICKNESS(M)" ;TH, "CONTACT RADIUS(M)" ;R1
581 PRINT "NO. OF CONTACTS" ;N
582 PRINT "OXIDATION TEMP,TO,(DEG.C)" ;TF
583 PRINT
584 PRINT
590 ON SGN(I)+(2) GOTO 600,620,640
600 LET I=I+1
601 LET APT=0.87E3
602 LET QP=96000
603 LET F=0.2885
604 LET RH=5.21E3
605 LET RD=2.0E-6
606 PRINT "IF FE3O4 THEN:-"
608 REM LINES 601 THROUGH 606 RELATE TO FE304 OXIDATION
610 GOTO 430
620 LET I=I+1
621 LET APT=2.86E7
622 LET QP=210000
623 LET F=0.2277
624 LET RH=5.7E3
625 LET RD=2.0E-6
626 PRINT "IF FEO THEN:-"
628 REM LINES 620 THROUGH 626 RELATE TO FEO OXIDATION
630 GOTO 430
640 NEXT K
650 STOP
660 PRINT "NO SOLUTION FOUND"
670 END

This article concludes with a summary of the way to use the oxidational wear model to analyze practical tribosystems and an appraisal of the future direction of oxidational wear research as it relates to the interests of typical design engineers or metallurgists who want to understand ad make practical provision in their tribosystem for the changes in geometry brought about by wear and frictional heating.

Mechanisms of Wear

The recent paper by Lim and Ashby (Ref 4) discusses the various mechanisms of wear, with special reference to the sliding wear of cylindrical steel pins against steel disks. After careful analysis of the results and conditions used by several tribological research groups, these authors summarize their findings by means of a wear mechanism map for steels. Figure 1(a) shows the boundaries between seizure, melt wear, severe oxidational wear, mild oxidational wear,

delamination wear, and ultramild wear. The abscissa is the normalized velocity. U_n, and the ordinate is the normalized pressure, W_n, defined by:

$$U_n = \frac{(U)(RT)}{XK} \quad \textbf{(Eq 1a)}$$

$$W_n = \frac{W}{(A_n)(PP)} \quad \textbf{(Eq 1b)}$$

where U is the linear velocity at the pin surface, W is the applied normal load, RT is the radius of the cylindrical pin, A_n is the nominal, or apparent, area of contact between the pin and the disk, XK is the thermal diffusive of the underlying metal, and PP is the room-temperature hardness of the underlying metal of the pin. The contours of constant normalized wear rate, WR_n, are superimposed on the field, thereby showing the regimes of dominance of the different wear mechanisms. The normalized wear rate is defined by:

$$WR_n = \frac{WR}{A_n} \quad \textbf{(Eq 1c)}$$

where WR is the wear rate (in units of volume removed per unit sliding distance).

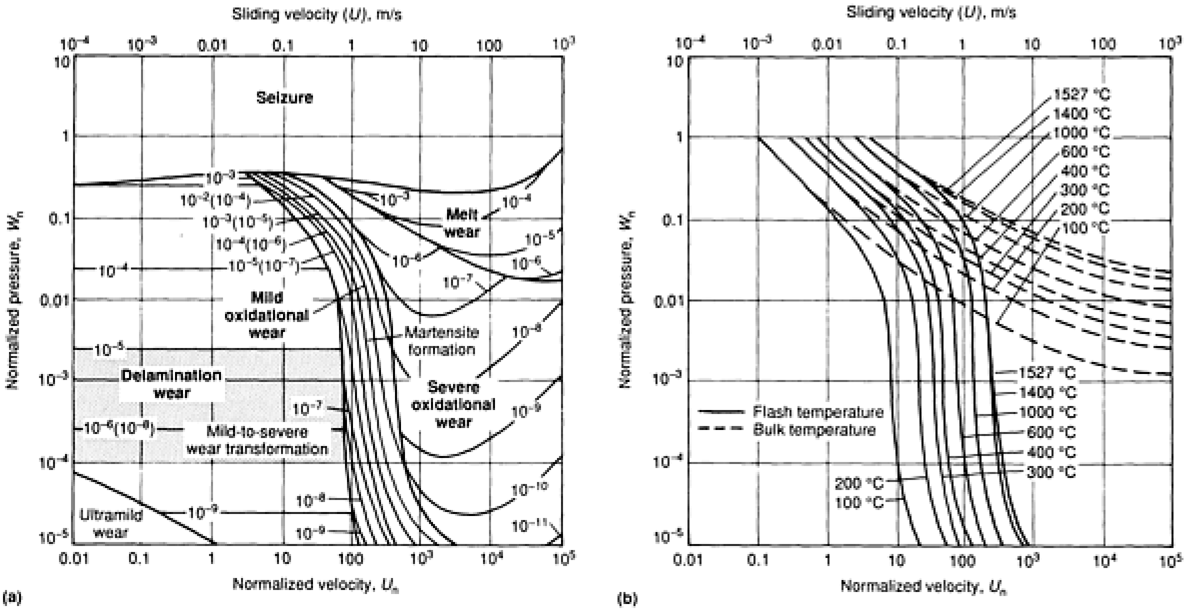


Fig. 1 (a) Wear mechanism map for steel sliding pairs using the pin-on-disk configuration. (b) Temperature map for steel sliding on steel in the pin-on-disk configuration

Figure 1(a) provides the design engineer with a useful guide to the wear rate to be expected in a steel-on-steel tribosystem, for given normalized velocity and normalized pressure. Lim and Ashby (Ref 4) have also produced another graph (Fig. 1b), which gives the designer some idea of the general surface temperature ad contact temperatures to be expected for given values of W_n and U_n.

It should be emphasized that these estimates are for a pin-on-disk system; are only very approximate, as far as the actual wear rates are concerned; and are somewhat lower than expected as far as contact temperatures are concerned (the surface temperatures are estimated to a much closer degree).

However, the particular strength of the wear mechanism map of Lim and Ashby is its all-inclusive nature, at least in terms of the wear of steels. For the first time, the work of the various tribological research groups has been brought together in a single schematic representation, which obviously will be useful for design engineers.

Mechanism of Mild Oxidational Wear

Our knowledge of the mechanism of mild oxidational wear has evolved gradually since the late 1960s (Ref 5, 6, 7, 8, 9) as a result of extensive microscopic and crystallographic analyses of worn surfaces and wear debris. In the initial stages of a wearing process, severe wear occurs. During this period, opposing surfaces achieve conformity, so that the real areas of contact consist of several large plateaus, the area of each being about the size of the real area of contact expected from the plastic deformation theory of Bowden and Tabor (Ref 10), namely *W/PP*, where *W* is the applied load and *PP* is the Brinell hardness (expressed in units of stress). At any given instant, one of these plateaus bears most of the load and becomes the site of considerable frictional heating. This causes the metal below the plateau surface to expand, in a manner similar to that described by Barber (Ref 11), and become the only region of contact until it is removed by wear. If the sliding speed is comparatively slow or the loads are so light that frictional heating is negligible, then the expansion of the contacting plateau will not be sufficiently large for it to become a preferred contact region. Furthermore, the rate of oxidation of the contacting plateau will not be very different from that of the remainder of the surface.

However, given sufficient frictional heating, the contacting plateau will oxidize preferentially, compared with the remainder of the surface. It will then oxidize at a temperature, *TF,* that is normally well in excess of the general surface temperature, *TS.* The surfaces of these plateaus tend to be extremely smooth, with fine wear tracks parallel to the direction of sliding. Typically, each plateau has an area of about 0.01 mm^2 (15.5 mil^2) and a height of about 3 or 4 μm (120 or 160 μin.).

The plateaus will sometimes exhibit cracks running at right angles to the sliding direction, somewhat similar to the fatigue crack systems found in fracture mechanics. It is possible, but not yet proven, that the intermittent heat and stress cycles suffered by the plateaus, as they enter and leave contact with their "opposite numbers," could bring about wear through the contact fatigue mechanism indicated by these crack systems. Typically, the surfaces surrounding these plateaus are rough and strewn with wear-debris fragments that are the remnants of previously existing contact plateaus.

The contacting plateau (on each opposing surface) is the site for all the asperity-asperity interactions between two sliding surfaces at any given moment. According to the mild oxidational wear mechanism, these asperities are the site for oxidation at the contact temperature. Because oxidation occurs by the diffusion of oxygen ions inward and, sometimes, by metal ions outward, one would expect the plateau to grow in height from the interface between the oxide and the metal beneath each asperity contact. In the course of many passes, it is reasonable to expect that the increases in height would be spread over the whole area of the contacting plateau. Upon reaching a critical oxide film thickness, the plateau becomes unstable and breaks up to form flakes and, eventually, wear debris.

The mild oxidational wear mechanism does not attempt to explain why the plateau breaks off at a critical thickness of a few microns. However, the theory does assume that when the contacting plateau finally breaks up, then another plateau elsewhere on the surface becomes the operative one. The virgin metal beneath the original plateau becomes free to oxidize at the general surface temperature. Without externally induced heating, the amount of oxidation at those parts of the surface that are not in contact (typically at a temperature of about 80 °C, or 175 °F) will be several orders of magnitude less than the oxide growth at a contacting plateau (at a typical temperature of 400 °C, or 750 °F).

Mild Oxidational Wear Theory

Any type of wear must involve the real are of contact, A_r, between surfaces in a state of sliding loaded contact. The original papers on mild oxidational wear began with the Archard wear law (Ref 12), which states that the wear rate is proportional to the real area of contact. This can be expressed mathematically as:

$$WR = (K)(A_r) = \frac{KW}{PP} \qquad \textbf{(Eq 2)}$$

where *K* is dimensionless constant and *PP* is the room-temperature hardness of the softer of the sliding pair.

Archard (Ref 12) interpreted this *K*-factor as the probability of producing, on the average, a wear fragment at each asperity encounter. This would mean that an average of 1/*K* encounters are needed to produce a wear fragment. Quinn (Ref 6) suggests that, for oxidational wear, this could also mean that 1/*K* encounters need to occur at a plateau in order for it to build up to the critical oxide thickness.

Using this interpretation, the following expression was evolved:

$$WR = \frac{2(A)(KP)(W)}{(TH)^2(RH)^2(F)^2(U)(PP)} \quad \textbf{(Eq 3)}$$

where *A* is the radius of each asperity contact, *RH* is the average density of the oxidized plateau, *F* is the fraction of oxide that is oxygen, and *KP* is the parabolic rate constant.

This constant depends on the temperature of oxidation through the expression:

$$KP = (AP)\exp\frac{-QP}{(R)(TF)} \quad \textbf{(Eq 4)}$$

where *AP* is the parabolic Arrhenius constant, *QP* is the parabolic activation energy, and *R* is the universal gas constant. Equation 3 can be written in terms of *N*, the number of asperity contacts, by inserting $A = \{(W)/[(\pi)(N)(PP)]\}^{1/2}$, thereby giving an equation for the wear rate in terms of: the surface model for oxidational wear (*N, TH,* and *TF*); various constants (*AP, QP, PP, RH, F, R,* and π); and the operating parameters (*W* and *U*).

More recent work by Hong, Hochman, and Quinn (Ref 1) uses an alternative derivation of the mild oxidational wear equation, which does not depend on the validity of the Archard (Ref 12) interpretation of the *K*-factor. The derivation is simply an equivalence between the total volume of wear and the volume of material (that is, oxide) formed on the real area of contact in the time required to build up a critical oxide film thickness, namely:

$$(A_r)(TH) = (WR)(U)(t) \quad \textbf{(Eq 5)}$$

One can now involve oxidation theory by assuming that the mass uptake of oxygen per unit area occurs either parabolically or linearly with respect to the time, *t.* The linear assumption produces the equation:

$$(RH)(F)(TH) = (KL)(t) \quad \textbf{(Eq 6)}$$

where *KL* is the linear oxidation rate constant. From these last two equations, and by replacing A_r with *W*/*PP,* an alternative expression for the wear rate is obtained:

$$WR = \frac{(KL)(W)}{(RH)(F)(U)(PP)} \quad \textbf{(Eq 7)}$$

where

$$KL = (AL)\exp\frac{-QL}{(R)(TF)} \quad \textbf{(Eq 8)}$$

and *AL* and *QL* are the linear Arrhenius and linear activation energies, respectively. Note that Eq 7 does not contain a term in *TH.* In fact, it is an expression with:

- One explicit surface model term, namely *TF*
- Several constants, namely *AL, QL, PP, RH, F,* and *R*

- The operating parameters, *W* and *U*

However, it should be remembered that *TF* is itself a function of both *N* and *TH*.

Whether one approaches the theory of mild oxidational wear from the Archard (Ref 12) interpretation of the *K*-factor or from the straight-forward equivalence between the volume of wear and the volume of oxide formed on the real areas of contact, one should end up with the same expression. It is a matter of pure algebra to show that both approaches give the same equations for linear oxidation, provided that one assumes the distance of a wearing contact, namely 2(*A*), is of the same order of magnitude as the thickness of the oxide. The same equivalence holds for parabolic oxidation.

Next, the heat flow aspects of oxidational wear are considered, because they lead to theoretical values for the division of heat, *DTHP,* and the contact temperature, *TF.*

Calculation of Heat Flow

Archard (Ref 13) calculated the heat flow between a moving and a stationary heat source, without taking into account the effect of oxides on both the pin and disk of his tribosystem. He also assumed that the general surface temperature of the pin was not different from that of the disk surface, so that he could apply the conventional thermal resistance analog for obtaining both the theoretical division of heat along the pin, *DTHP,* and the theoretical temperature excess, *TE,* at the real areas of contact over the general surface temperature (*TSP* or *TSD*) of the pin or disk, respectively, namely:

$$DTHP = \frac{TD}{(TD) + (TP)} \qquad \textbf{(Eq 9)}$$

$$TE = \frac{(TP)(TD)}{(TD) + (TP)} \qquad \textbf{(Eq 10)}$$

In these equations, *TE* is equal to (*TF* - *TSP*), *TD* is the fictitious excess temperature of the disk contact areas, assuming (for the sake of discussion) that all of the heat generated at the interface passes into the disk, and *TP* is the fictitious excess temperature of a pin asperity, assuming (again, for the sake of discussion) that all of the frictionally evolved heat passes into the pin.

A later modification of the Archard (Ref 13) expressions has been published by Quinn, Rowson, and Sullivan (Ref 2). These authors have included the oxide thickness on the pin, *TH,* in the expression for the theoretical division of heat along the pin, $DTHP_{qrs}$, and the theoretical contact temperature, TF_{qrs}, namely:

$$DTHP_{qrs} = \frac{(TD)_{qrs}}{(TD)_{qrs} + (TP)_{qrs}} \qquad \textbf{(Eq 11)}$$

where

$$(TP)_{qrs} = \frac{HT}{4(A)(KS)(N)} + \frac{(HT)(TH)}{(\pi)(A^2)(KO)(N)} \qquad \textbf{(Eq 12)}$$

and TD_{qrs} is given by one of two relationships, depending on whether *L* is greater or less than 5, where *L* is given by:

$$L = \frac{(U)(A)}{2(XK)} \qquad \textbf{(Eq 13)}$$

If $L > 5$, then TD_{qrs} is given by TDG, where:

$$TDG = \frac{0.85\ HT}{4(A)(N)(KS)\sqrt{L}}$$ **(Eq 14)**

If $L = 5$ or $L < 5$, then TD_{qrs} is given by *TDL,* where:

$$TDL = (-0.1021L + 0.8605)\left[\frac{HT}{4(A)(N)(KS)}\right]$$ **(Eq 15)**

The modified expression for the contact temperature, TF_{qrs}, can be written as:

$$TF_{qrs} = [(DTH)_{qrs}\ (TP)_{qrs}] + TSP$$ **(Eq 16)**

Some new variables have been introduced in the last few equations, namely *KO,* the thermal conductivity of the oxide, and *HT,* the total heat evolved through friction at the sliding interface, which is equal to [(*U*)(*FF*)]. With these theoretical heat-flow equations, and the mild oxidational wear equations given previously, theory can now be compared with the results from selected experiments with a low-alloy steel.

Theory and Experimental Results Comparison

In this section, the experimental values of wear rate and division of heat are used, together with the FINDAP computer program (details of which are given in Table 1), to deduce tribological oxidation constants that enable the use of theoretical expressions for the division of heat and the wear rate for similar tribosystems. Also used in this section are the values of the tribological oxidation constants described above, together with the OXYWEAR computer program (details of which appear in Table 2), to deduce consistent values of the radius of asperity contact, *A.* These values of *A* are, in turn, used to generate consistent values of the number of contacts, *N,* the critical oxide thickness, *TH,* and the contact temperature, *TF.*

First described are the relevant tables of data. Table 3 contains data for a series of experiments carried out with En8 steel at a speed of 2 m/s (6.5 ft/s), using a pin-on-disk wear machine that was fully instrumented for measuring the division of heat along the pin. This table was extracted from Table 1 of Ref 2. The symbols *TA, TB, TC,* and *H*1 relate to the thermocouple measurements and the heat flow calculated from these measurements for the pin-holder, as shown in Fig. 2.

Table 3 Data for a series of wear tests with En8 steels

Parameter	Experiment number				
	1	**2**	**3**	**4**	**5**
Speed, m/s (ft/s)	1.97 (6.27)	1.97 (6.27)	1.97 (6.27)	1.97 (6.27)	1.97 (6.27)
Load, N (lbf)	3.94 (0.88)	9.85 (2.21)	19.70 (4.41)	29.55 (6.62)	39.4 (8.83)
Friction, N (lbf)	4.66 (1.04)	9.07 (2.04)	14.22 (3.18)	13.98 (3.13)	17.66 (3.96)
***TA*, °C (°F)**	37.1 (99)	47.0 (117)	67.2 (135)	74.0 (165)	149.0 (300)
***TB*, °C (°F)**	30.8 (87)	35.8 (96)	46.5 (116)	53.0 (127)	94.0 (201)
***TC*, °C (°F)**	27.5 (82)	28.7 (84)	34.2 (94)	33.0 (91)	48.0 (118)
***H*1, W**	0.50	0.90	1.65	1.78	4.58
***HT*, W**	9.18	17.87	28.01	27.54	34.79
DE	0.0542	0.0505	0.0591	0.0648	0.1318
***TS*, °C (°F)**	39.78 (103.6)	51.85 (125.3)	76.10 (169)	83.54 (182.4)	173.55 (344.4)
WR	2.063	4.030	3.003	4.418	17.890

WR, wear rate in units of $m^2 \times 10^{-13}$; *HT,* total frictional heat at the sliding interface; *DE,* fraction of *HT* passing along the pin; *TS,*

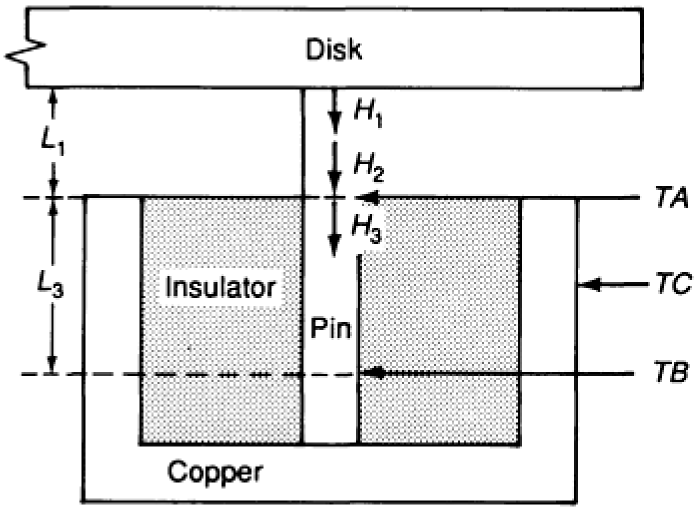


Fig. 2 Heat flow diagram for a pin loaded against a disk rotating about an axis parallel to the pin, but situated some distance to the left of the diagram

Table 4 relates to the results obtained using the FINDAP computer program, in conjunction with the data in Table 3. The first few rows of Table 4 give the main constituents of the wear debris, as identified by x-ray diffraction. This identification enables one to select the appropriate oxidational activation energy, because Fe_3O_4 cannot appear in the wear debris unless the contact temperature, *TF,* is over 450 °C (840 °F), and FeO can only appear for *TF* values that are greater than 600 °C (1110 °F). This table shows that the tribological Arrhenius constants, *APT,* for this particular tribosystem are several orders more than the Arrhenius constants, *AP,* obtained from Caplan and Cohen (Ref 14).

Table 4 Using FINDAP on results of Table 1

Parameter	Experiment number				
	1	**2**	**3**	**4**	**5**
Wear debris analysis	Only Fe_2O_3	Only Fe_2O_3	Fe_2O_3 and Fe_3O_4	Fe_2O_3 and Fe_3O_4	Fe_2O_3, Fe_3O_4, and FeO
DE	0.05416	0.05045	0.05906	0.0648	0.1318
DTH	0.05416	0.05045	0.05906	0.0648	0.1318
***TH,* μm (μin.)**	2.0 (80)	2.5 (100)	3.25 (130)	4.25 (170)	2.50 (100)
N	398	546	195	57	12
***TF,* °C (°F)**	440 (825)	415 (780)	518 (964)	512 (954)	769 (1415)
***A,* μm (μin.)**	0.98 (39)	1.33 (53.2)	3.28 (131)	7.2 (290)	18.0 (720)
APT	2.92E12	9.33E12	0.98E3	0.77E3	2.86E6
APT	1.5×10^6	1.5×10^6	0.032	0.032	1.1×10^5
***QP,* × 10^3**	208	208	96	96	210

DE, experimental division of heat along the pin; *DTH,* theoretical division of heat along the pin; *TH,* thickness of oxide film on contact plateaus; *N,* number of asperities in contact beneath the pin; *TF,* contact temperature; *A,* radius of an individual asperity contact; *APT,* tribological Arrhenius constant; *AP,* Arrhenius constant for static oxidation (Ref 2); *QP,* oxidational activation energy

Table 5 relates to data obtained from a series of wear tests with another low-alloy steel, En31, sliding against itself at a speed of 4 m/s (13 ft/s). This table is particularly interesting, because it gives the actual proportions of the various oxides in the debris. Later, it will be shown how this information can be used to overcome the problem of two or three different oxides of the metal being sandwiched within each of the contacting plateaus. Note that the debris for Experiment 1 was entirely alpha iron, indicating that this was not oxidational wear, so that the oxide thickness would be virtually zero. For the sake of later comparison, the variation of wear rate, *WR,* general surface temperature, *TS,* and the experimental division of heat, *DE,* with load, *W,* is shown in Fig. 3(a-c)

Table 5 Data for series of wear tests with En31 steels sliding against each other at a linear speed of 4 m/s (13 ft/s)

Parameter	Experiment number				
	1	**2**	**3**	**4**	**5**
Speed, m/s (ft/s)	4.00 (13.1)	4.07 (13.4)	4.07 (13.4)	4.07 (13.4)	4.07 (13.4)
Load, N (lbf)	3.94 (0.88)	9.85 (2.21)	29.55 (6.62)	39.40 (8.83)	54.17 (12.13)
Friction, N (lbf)	4.41 (0.99)	5.40 (1.21)	13.05 (2.92)	19.38 (4.34)	25.44 (5.70)
***TA*, °C (°F)**	30.8 (87)	50.0 (122)	86.0 (187)	112.5 (235)	122.0 (252)
***TB*, °C (°F)**	25.2 (77)	37.1 (99)	56.5 (134)	52.2 (126)	87.0 (189)
***TC*, °C (°F)**	22.0 (72)	28.1 (83)	34.8 (95)	42.5 (109)	52.0 (126)
***H*1, W**	0.458	1.081	2.49	4.55	3.09
***HT*, °C (°F)**	17.44 (63.4)	21.98 (71.6)	53.11 (127.6)	78.88 (174.0)	103.54 (218.4)
DE	0.026	0.0492	0.0468	0.0577	0.0299
***TS*, °C (°F)**	33.23 (91.8)	55.72 (133.3)	99.12 (210.4)	137.0 (278.6)	138.27 (280.9)
***WR*, $m^2 \times 10^{-13}$**	0.335	1.357	2.769	3.119	6.172
Proportional analysis of wear debris[(a)]					
Fe, %	100	8	1	0	1
Fe_2O_3, %	0	68	49	37	30
Fe_3O_4, %	0	24	50	63	69
FeO, %	0	0	0	0	0

(a) Proportions obtained from x-ray diffraction analysis

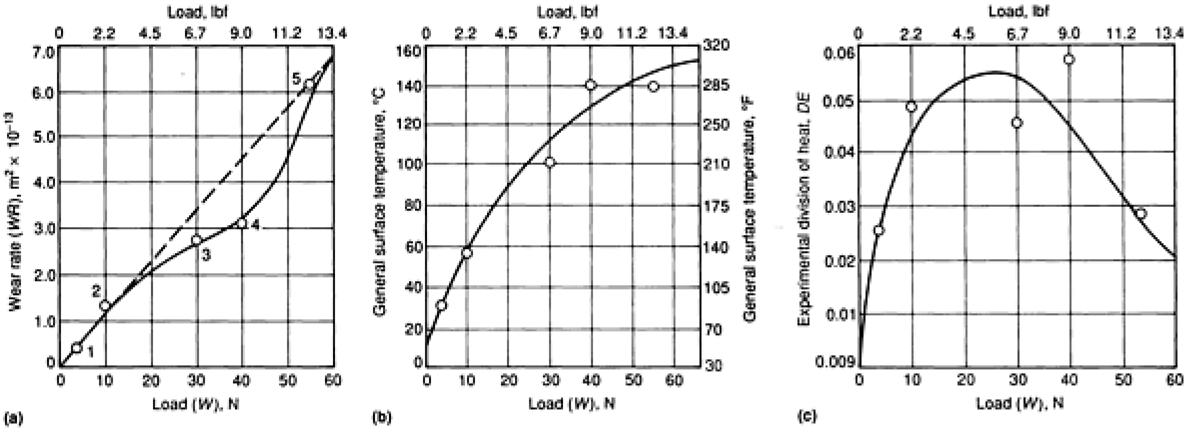


Fig. 3 Variation with load, *W,* of (a) Wear rate, *WR.* (b) General surface temperature, *TS.* (c) Division of heat along the pin, *DE,* for En31 steel sliding at 4 m/s (13 ft/s)

The wear rate versus load curve deviates from the expected linear shape for the experiments at 29.5 N (6.60 lbf) and 29.40 N (6.59 lbf), whereas the division of heat versus load curve shows an unexpected decrease in and reversal of slope at loads above approximately 20 N (4.5 lbf). The smooth shape of the general surface temperature versus load curve indicates that *TS* is unlikely to be connected with the deviant behavior of *DE* and *WR.*

If the general surface temperature is not the primary factor, perhaps the effect of changes in the type of oxide proportions should be considered, because it is well-known that Fe_3O_4 is a low-wear and low-friction material, compared with Fe_2O_3. In order to do this, the OXYWEAR computer program must first be used to deduce the values for *N, TH,* and *TF* for 100% Fe_2O_3, 100% Fe_3O_4, and 100% FeO. The results of this analysis are shown in Table 6.

Table 6 Expected values of *N, TH,* and *TF* based on the OXYWEAR computer program applied to the results of Table 3

Experiment number	At 100% Fe_2O_3			At 100% Fe_3O_4			At 100% FeO		
	N	***TH***	***TF***	***N***	***TH***	***TF***	***N***	***TH***	***TF***
1	181	1.95	396	97	2.62	523	44	3.80	743
2	305	2.07	387	150	2.83	495	50	4.43	722
3	219	2.72	387	105	3.70	492	31	5.74	716
4	786	1.20	372	422	1.60	450	116	2.81	692
5	778	2.90	406	286	4.56	560	110	6.80	769

For 100% Fe_2O_3 oxide plateaus, it is assumed that $APT = 6.12 \times 10^{12}$ and $QP = 208 \times 10^3$. For 100% Fe_3O_4, it is assumed that $APT = 0.87 \times 10^3$ and $QP = 96 \times 10^3$. For 100% FeO, it is assumed that $APT = 2.86 \times 10^7$ and $QP = 210 \times 10^3$.

Table 6 was compiled using the tribological Arrhenius constants obtained from the analysis of En8 steel applied to the results given in Table 5, assuming the three extreme cases described in the preceding paragraph. The OXYWEAR computer program, which was used to obtain these values, has already been published (in FORTRAN) (Ref 2). It uses a different approach to that used in the FINDAP program. In order to deduce values of *N, TH,* and *TF* that are appropriate to the experiments detailed in Table 5, the predictions of Table 6 must first be modified to reflect the fact that the wear debris for wear experiments 2 through 5 indicates that the contact plateaus must have oxidized at temperatures above 450 °C (840 °F), which is the temperature at which Fe_3O_4 begins to appear in the wear debris, but below 600 °C (1110 °F), which is the temperature at which FeO begins to appear in the wear debris. This modification is provided in Table 7, which has been constructed on the assumption that *N, TH,* and *TF* are linearly dependent on the proportion of Fe_3O_4 in the wear debris.

Table 7 Values of *N, TH,* and *TF* calculated from Table 4, assuming parameters are linearly dependent on proportion of Fe_3O_4 in wear debris

Experiment Number	Fe_3O_4, %	*N*	*TH*		*TF*	
			μm	**μin.**	**°C**	**°F**
1	0(a)	. . .	0(a)	0(a)	. . .	. . .
2	24	268	2.25	90.0	413	775
3	50	162	3.21	128	440	825
4	63	557	3.18	127	421	800
5	69	439	4.05	162	512	954

(a) Alpha iron was the only identified constituent of wear debris; hence, *TH* is probably close to zero.

Figure 4 illustrates how *TF, TH,* and *N* vary with the proportion of Fe_3O_4 in the debris, thereby permitting a ready comparison between these important tribological parameters. The curves in Fig. 4 are similar in shape to those of the wear rate versus load curve in Fig. 3(a), which shows that contact temperature, oxide thickness, percentage Fe_3O_4, and the wear rate are all directly related to one another. The smooth increase in the *N* curve is a mechanical effect of the increase in load, as could be shown by plotting *N* versus *W.*

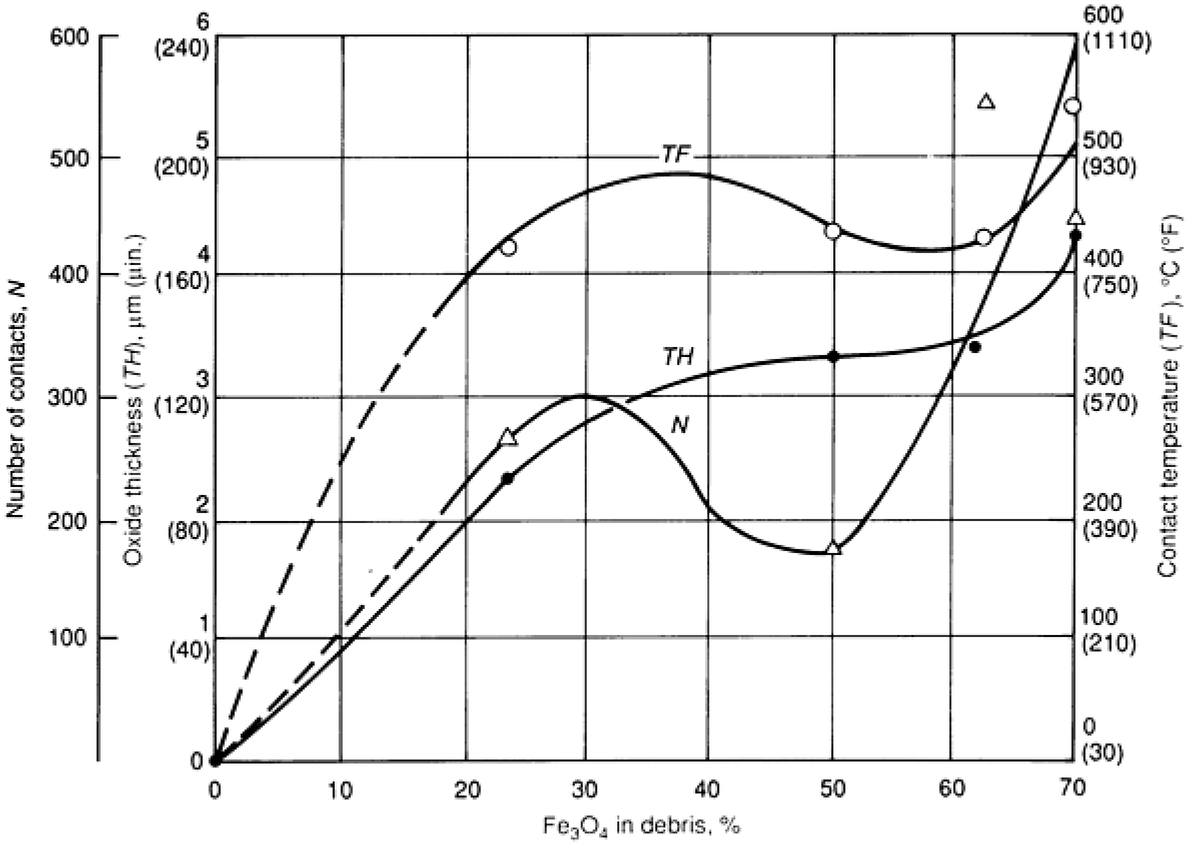


Fig. 4 Variation of oxide thickness, *TH,* number of contacts, *N,* and contact temperature, *TF,* with percentage of Fe_3O_4 in the wear debris from the wear experiments carried out with En31 steel sliding at 4 m/s (13 ft/s)

Although the model for oxidational wear and division of heat has been shown to yield interesting estimates of contact temperatures, oxide thicknesses, and numbers of asperity contacts, all of which can be shown to be of the right order of magnitude compared with values measured by electron microscopy, x-ray diffraction, infrared microscopy, thermometry, and heat-flow analysis, the model has yet to be applied to practical tribosystems. The next section in this article describes how the wear mechanism and temperature maps of Lim and Ashby (Ref 4) can be used, along with the ideas and results described in this review, to design a tribosystem that allows for the:

- Removal of oxidized surface layers from the triboelements over the proposed lifetime of the tribosystem
- Thermal expansion of the triboelements at *TS*
- Possible problems that could arise from very high contact temperatures

Application to Practical Tribosystems

The focus of the discussion below only concerns steel tribosystems that slide without any obvious lubrication (apart from the formation of oxidized surfaces). Nevertheless, there seems to be no reason why, in principle, the ideas and results discussed in this review should not be applicable to other metals that produce oxides under normal atmospheric conditions. Unfortunately, from a scan of the published literature, it appears that no metal other than iron or its alloys has received sufficient attention for it to be possible to devise a wear mechanism map similar to the one produced by Lim and Ashby (Ref 4). However, oxidational wear concepts have been applied to other metals, such as titanium, nickel alloys, and molybdenum. A broad program of research will be necessary to produce wear mechanism maps that are as all-inclusive as the steel wear mechanism maps for these metals and their alloys.

Assuming, therefore, that a particular tribosystem employs steels for its triboelements, one must first identify the wear mechanism and the relevant surface temperatures from Fig. 1(a) and 1(b). The normalized pressure, W_n, should not be

difficult to deduce, because it depends on the pressure and the room-temperature hardnesses of the triboelements. The normalized velocity, U_n, depends on the radius of the pin, as well as the thermal diffusivity of the disk. Clearly, the radius of the pin is related to the length of the area of nominal contact. Therefore, for a geometry different from that of a pin-on-disk, it is suggested that a characteristic length associated with the nominal area of contact of the stationary triboelement be used.

With these values of U_n and W_n, one can estimate the type of wear mechanism from Fig. 1(a), as well as the *TS* and the *TF* from Fig. 1(b). If the wear mechanism indicated by reference to these figures is mild oxidation wear, then one should be able to deduce the wear rate from Eq 3 and 4 using:

- The above estimates for *TF* and *TS*
- The values of *APT, QP, TH, RH,* and *F* given in this review for the relevant temperature ranges of the three oxides of iron
- The value of the load, *W,* from the pressure
- The value of the radius of asperity contact, *A,* taken to be approximately the same magnitude as the oxide thickness

Remembering that the wear rate is in units of volume removed per unit sliding distance, one can readily deduce a value for the thickness, or mass removed per unit distance of sliding or per unit time, for a particular tribosystem. Knowledge of the expected wear and the expected surface temperatures will help in the design of steel tribosystems to closer limits than heretofore, as well as avoid the use of materials, such as some solid lubricants, that will not function properly at high contact temperatures.

Future Trends

Although some research has been carried out on the connection between boundary lubrication and mild oxidational wear (Ref 15), this work has not been followed up. There is a clear need to take a similar approach toward understanding the competing chemical reactions of both the oxygen and the extreme-pressure additive in the oil with the wearing metal surfaces in a heavily loaded tribosystem. This situation calls for the application of the severe oxidational wear theory.

The theoretical aspects of severe oxidational wear have been known for some time (Ref 16, 17), but the theory still awaits experimental confirmation. However, it is very probable that, during the mid-1990s, the computer modeling methods described earlier will be successfully extended and applied to wear experiments carried out under elevated-temperature conditions. As soon as this modeling has resolved the problems involved in allowing for "out-of-contact" oxidation (Ref 18), the extension to heavily loaded tribosystems lubricated with extreme-pressure lubricants should be just a matter of time.

References

1. H. Hong, R.F. Hochman, and T.F.J. Quinn, A New Approach to the Oxidational Theory of Mild Wear, *STLE Trans.,* Vol 31, 1988, p 71
2. T.F.J. Quinn, D.M. Rowson, and J.L. Sullivan, The Application of the Oxidational Theory of Mild Wear to the Sliding Wear of Low-alloy Steel, *Wear,* Vol 65, 1980, p 1
3. J.L. Sullivan, T.F.J. Quinn, and D.M. Rowson, Developments in the Oxidational Theory of Mild Wear, *Tribol. Int.,* Vol 12, 1980, p 153
4. S.C. Lim and M.F. Ashby, Wear-Mechanism Maps, *Acta Metall.,* Vol 35, 1987, p 1
5. J.L. Sullivan and S.S. Athwal, The Mild Wear of a Low-alloy Steel at Temperatures up to 500 °C, *Tribol. Int.,* Vol 16, 1983, p 123
6. T.F.J. Quinn, The Effect of "Hot-spot" Temperatures on the Unlubricated Wear of Steel, *ASLE Trans.,* Vol 10, 1967, p 158
7. T.F.J. Quinn, The Dry Wear of Steel as Revealed by Electron Microscopy and X-ray Diffraction, *Proc. Inst. Mech. Eng.,* Vol 182, Pt3N, 1968

8. T.F.J. Quinn, An Experimental Study of the Thermal Aspects of Sliding and Their Relation to the Unlubricated Wear of Steel, *Proc. Inst. Mech. Eng.,* Vol 183, Part 3P, 1969, p 129
9. T.F.J. Quinn, The Division of Heat and Surface Temperatures at Sliding Steel Interfaces and Their Relation to Oxidational Wear, *ASLE Trans.,* Vol 21, 1978, p 78
10. F.P. Bowden and D. Tabor, *The Friction and Lubrication of Solids, Part 1,* Clarendon Press, Oxford, 1954
11. J.R. Barber, The Influence of Thermal Expansion on the Friction and Wear Process, *Wear,* Vol 10, 1967, p 155
12. J.F. Archard, Single Contacts and Multiple Encounters, *J. Appl. Phys.,* Vol 32, 1961, p 1420
13. J.F. Archard, The Temperature of Rubbing Surfaces, *Wear,* Vol 2, 1959, p 438
14. D. Caplan and M. Cohen, The Effect of Cold Work on the Oxidation of Iron from 100 to 500 °C, *Corros. Sci.,* Vol 6, 1966, p 321
15. J.L. Sullivan, Boundary Lubrication and Oxidational Wear, *J. Phys. D: Appl. Phys.,* Vol 19, 1986, p 1999
16. T.F.J. Quinn, Review of Oxidational Wear--Part II: Recent Developments and Future Trends in Oxidational Wear Research, *Tribol. Int.,* Vol 16 1983, p 305
17. T.F.J. Quinn, Review of Oxidational Wear--Part I: The Origins of Oxidational Wear, *Tribol. Int.,* Vol 16, 1983, p 257
18. T.F.J. Quinn, The Effects of "Out-of-Contact" Oxidation on the Oxidational Wear of Steels, *Proc. 3rd European Tribology Congress,* Warsaw, Poland, 1981

Surface Examination

K.C. Ludema, University of Michigan

Introduction

TRIBOLOGICAL SYSTEMS require extensive examination of their surfaces to isolate potential sources of wear. The systems are classified as "tribological" systems in order to encompass problems that include wear by sliding, wear by erosion, chemically enhanced loss of material, friction without significant wear, and any other mechanical interaction of two substances, whether lubricated or unlubricated.

Surface examination may require the use of several types of instruments. Generally, tribologists can usually solve two-thirds of their problems using a small magnet, a low-power optical microscope or hand-held lens, and surface roughness tracing. Few tribologists will need the more sophisticated instruments, and even fewer can be expected to know how to operate them.

Little is written on methods of surface analysis for tribological problems. Analysis involves human decision as well as instruments. The best method of analyzing surfaces begins with a good plan, and the plan should include several steps. In the following discussion, it will be assumed that a problem is known to exist. Perhaps a candidate material is operating in some new device and some judgment must be made of its suitability. Perhaps some surfaces are wearing too quickly or in some undesirable pattern, or the surfaces may be sliding in some undesirable manner, and the time has arrived to examine those surfaces. A procedure for surface analysis is given in the following paragraphs. The very first, and perhaps surprising suggestion is to *avoid dismantling the device or cleaning the surfaces* before performing the steps outlined below, either formally or informally.

Planning

Assemble a group of people (depending on the size of the problem) consisting of:

- Engineers and technicians who have responsibility for the product under discussion. Wear is influenced by the system surrounding a set of sliding surfaces in addition to the material composition of the sliding surfaces, and several skills should be brought into the discussion
- One or more persons with several years experience in general problems in friction, lubrication, and wear. These people serve as valuable buffers between product engineers,who need a "quick fix," and instrument specialists, who prefer to be more thorough
- Specialists in solid mechanics,fluid mechanics, lubricant chemistry, materials science, physics, and so on, should be consulted. These specialists must be selected with care, particularly if they are remote from practical problems. Surface scientists in particular, tend to concentrate on very fine detail, which *seems* sensible, but may not be. Their expertise is vital, however, and can best be applied when problems can be "broken down" into workable segments by people with broader experience in tribology

Develop a case history to gain a perspective on how the impressions or convictions were developed that the surfaces in question were judged to be operating either properly or improperly. Determine the conditions under which the surfaces seem to behave improperly. If the undesirable phenomenon appears periodically, determine whether this behavior is related, for example, to a change in supplier, a change in weather, a change in the "observer," or a change in the process sequence for making the original surface.

Develop a suitable expression for wear rate or performance problems of the surfaces in question. Are the surfaces wearing progressively, are they scuffing, is there vibration at certain times and not at others, and so on? Can these phenomena be quantified?

Decide between examining the wearing surfaces themselves or measuring the effect of wear (or uneven friction, and so on) on the functioning of the machine or component in question. It may be easier or more economical to redesign a machine component to accommodate a particular wear rate or frictional behavior than to find new materials to reduce wear rate or provide more predictable friction. Perhaps both will be necessary. The measurement of component function will probably involve measurement of changes in part clearances, friction, vibration mode, and so on. Accommodating a given friction or wear rate is a design question, which will not be discussed further.

If surface examination is necessary, it is useful to plan the steps leading to such examination, as discussed in the following sections of this article.

First Level of Surface Examination

Check Effect of Mechanical Test Sequence on Surface Chemistry. Determine, if possible, what effect there will be on the surfaces in question by stopping sliding (eroding, and so on) by dismantling the mechanical system containing the surfaces and by cleaning the surfaces. In some instances the surface chemistry will change with time after the machine is shut off, and surface chemistry will surely change during cleaning. In many instances a test device cannot be stopped, taken apart for examination, and reassembled without making some undesired change.

It is often important to preserve the wear debris on and near sliding surfaces for analysis. Observe the location of the buildup of debris, the flow patterns of debris, the particle size distribution, and so on. Obtain an oil sample and save any filters in the system that are replaced.

Dismantle the mechanical system in question in the presence of the personnel responsible for its performance. Note the practices of the persons doing the dismantling, and the possible effect of their practices on the surface condition of the workpiece.

Conduct a Preliminary Investigation by Human Senses. Use eyes, fingers, and nose to make a first judgment of the environment in which the surfaces are operating. There may be "gritty" substances or ridges of debris on or near the sliding surfaces, or there may be some particular pattern of marks, pits, or plowed ridges on the surfaces. A 10× eyepiece (magnifying glass) is probably the best aid at this stage.

Remove the surfaces to be examined and obtain some wear debris.

Observe the surfaces and debris in a binocular microscope that has a magnification range from about 2 to 40× (see the article "Optical Microscopy" in this Volume). Use a movable light source to light the target at all angles, from near-vertical to near-grazing angles. Rotate the specimens under the microscope to observe directional features of the surface.

Analyze Condition of Workpiece Surface. Surface materials may be worn away, rearranged, or built up by transfer. A perspective on these events can often be gained by surface tracing with a surface roughness tracer system (see the article "Wear Measurement" in this Volume) or other method of recording surface topography. Weighing of tribological components is sometimes useful. An important point is that the measurement of volume loss (or gain, as by material transfer) alone by any of the available methods is not sufficient. The shape of the worn region, the direction of scratching, the distribution of built-up material, and so on, must all be noted.

Repeat the above six steps for several specimens obtained from mechanical systems operated in various ways, with several different materials, and with different surface conditions, until every observer is sure of the sequence of surface change that is occurring and all agree on the scale of observation needed for full understanding of what is taking place. (See the section "Matters of Scale" in this article.)

It is important (albeit difficult) to obtain specimens in various stages of wear. When a tribological problem first appears, most investigators become very well acquainted with the failed state of the surfaces. However, before the final state, the surfaces had probably gone through several stages of change. The solution to problems often involves preventing the first stage of wear or change in behavior.

Proceed with patience. Interesting details of the debris and sliding surfaces are usually not obvious in the first hour of study, but with practice the eye eventually "sees" differences.

Develop a Hypothesis Concerning Surface Performance. The best hypothesis will arise from a group of people with the widest knowledge of tribological mechanisms. The hypothesis may contain elements that suggest the need for further analysis of some parts of the system, perhaps by an outside expert or vendor. For example, it might be postulated that the problem arises from vibration, may involve micropitting or hydrogen embrittlement (if the material is hardened steel), or may involve the buildup of compacted debris or chemical compounds from a lubricant.

Proceed with Laboratory Analysis or Conduct Modified Tests. With the hypotheses developed in the previous step, a choice may now be made between proceeding with laboratory analysis or proceeding with the further testing of practical parts. In most instances, further microscopic or chemical examination will not be as useful as empirically altering some part of the sliding system (for example, the materials, assembly practices, lubricants , and so on). However, if further examination is necessary, proceed to the next section.

Second Level of Surface Observation: Electron Microscopy

Scanning electron microscopy (SEM) is probably the most useful secondary analytical tool for surface analysis in tribology (see the article "Electron Microscopy" in this Volume). Most SEM units can cover a range of magnification from 20× to more than 30,000×. One major precaution in the use of the SEM system is that an effort must be made to retain perspective of size and scale. Perspective may be lost for two reasons. First, scanning electron microscopy has a depth of field that is ~300 times larger than that of the optical microscope at high magnification. This provides the advantage that most details on a rough surface will be visible, but it has the disadvantage that surfaces appear to be very much smoother than they actually are. Second, the great temptation in using SEM equipment is to focus on details that appear interesting, but that often turn out to be irrelevant.

Specimens for scanning electron microscopy must usually be small, typically $\leq$20 mm ($\leq$0.8 in.) thick and $\leq$80 mm ($\leq 3\frac{1}{8}$ in.) in diameter, depending upon the particular brand of SEM apparatus. They must be cleaned of volatile substances (unless the specimen can be cooled to cryogenic temperature in the scanning electron microscope) because the specimen will be inserted into a vacuum of better than 1.3 mPa (10^{-5} torr). If the specimen is a nonconducting material, it must be coated with carbon or gold so that an electron charge does not build up on the surface and deflect incoming electrons.

Images obtained with the SEM instrument do not correspond exactly with what is seen with the optical microscope. Scanning electron microscopy produces an images because the polarities across the specimen surface vary slightly.

Regions of positive bias appear dark and regions of negative bias (or with accumulated negative charge) appear bright. In contrast, the optical microscope produces an image of contrasting light intensities.

SEM units are often equipped with energy-dispersive x-ray analysis (EDAX) instrumentation (see the article "X-Ray Characterization of Surface Wear" in this Volume) for the purpose of identifying atomic elements in chosen regions on surfaces. In the most modern and automated instruments, interesting regions can be brought into the field of view, small details can be outlined within that field of view, and the elemental composition of the surface material within those outlines can be printed out directly.

Sometimes totally unexpected elements will appear in the analysis which do not relate to the specimen being examined. This occurs most often when scattered electrons in the specimen chamber impinge on the specimen holder or some other part of the instrument in the vicinity of the specimen, or it may due to a partially obstructed electron column.

The operation of modern SEM units equipped with EDAX instrumentation does not require high skill. However, skilled operator should be available to clean, align, and calibrate the instrument on occasion and to aid in the interpretation of some results.

Transmission electron microscopy (TEM) is a second type of electron microscopy (see the article "Electron Microscopy" in this Volume). It provides a view through a thin layer of solid material of ≤100 nm (≤1000 Å) thickness, depending on the voltage of the electron beam. Specimen preparation requires skill and patience because it is usually done by etching away unwanted material (usually with chemicals). Surface features of specimens can be observed with transmission electron microscopy, but this requires the making of replicas, shadowing, and several other time-consuming steps. TEM units have high-resolution capability, but require skilled personnel to operate the unit. Modern TEM instruments may also be equipped with electron diffraction instrumentation, which has several advantages over x-ray diffraction.

Selection of Chemical Analysis Instruments

The following steps must be taken to obtain tangible results using chemical analysis.

Isolate Specific Parameters to Be Investigated. Decide what information is desired from the surfaces under examination. This is necessary in order to choose the proper type(s) of instruments and to avoid a deluge of costly but superfluous information. The type of information needed may include the following (additional information is available in the section "Analysis of Data Collected"):

- Integrity of the original materials as indicated by surface cracks, loose grains, residual stresses, unexpected phases, inclusions, laps, and folds from surface processing, and so on
- Chemistry of "used" surfaces. Oxides, sulfides, organic compounds, decomposed lubricants, foreign matter, mixtures of phases from the original substrate materials, and so on, all influence sliding and wearing performance of machine components

Match Data Requirements with Instrument Capabilities and Limitations. Compare the capability of instrumentation with information needed from the instruments. This includes:

- The scale of depth and width. No practical sliding system that consists of material *x* sliding against, or eroded by, material *y* remains in its original state. After a short time each material is coated with other chemical species. If the coating is very thin (for example, 10 nm, or 100 Å) then the analysis of that coating must be done with an instrument that "penetrates" no deeper than the coating. On the other end of the scale, if the coating is thick (for example, 100 nm, or 1000 Å) and its composition varies throughout its thickness and over its expanse, then one instrument reading from the top three atomic layers over a target diameter of 10 atoms will provide data of very limited value
- Some analytical instruments identify elements only, and others provide information from which candidate compounds can be inferred. Most instruments operate within a limited range of the periodic table of the elements, but are, ironically, unable to identify the most common elements (namely,

hydrogen, carbon, and to a lesser extent, oxygen) on sliding surfaces. Time requirements, instrument charges, and operator expertise are usually proportional to the amount of information available from the instruments (as well as time required for the analysis)

- Several instruments operate with specimens confined in a vacuum. Volatile substances are usually not allowed into these instruments by the operator unless provision is made to cool the materials to a very low temperature. Nonvolatile substrates are usually rigorously cleaned before being placed into the vacuum. Unfortunately, this cleaning removes many of the coatings of interest

Define and Interpret Data to Provide Solution to Failure Analysis Problem. Develop a statement of precisely what data are to be obtained and how to interpret the results in a manner that is useful to the examination exercise. Instrument operators can explain results in terms of elements and compounds but not always in terms that are useful for solving a failure analysis problem.

Additional information about chemical analysis is contained in the article "Surface Chemical Analysis" in this Volume.

Analysis of Data Collected

The collected information must be studied with the intention of solving the problem. This seems obvious and trite to state, but the point is often forgotten in the information-gathering stage. It is often very tempting to explore in more and more detail. In most cases, it is better to "stand back" from the collected information and review the reason for the investigation underway. Check again to see whether or not the collected information answers questions on a macroscopic scale rather than on a microscopic scale. Check again to see whether or not some surviving parts (surfaces) have been investigated alongside the failed or failing parts. The first priority must be to solve the problem, the second to improve the product. Of course, an additional consideration might be to learn enough to publish a paper on the subject.

Matters of Scale

Size Scale of Objects. When examining specimens under a microscope, it is often useful to think about the scale of observation relative to the scale of size of various things. Figure 1 shows such a scale marked in SI units and in English units.

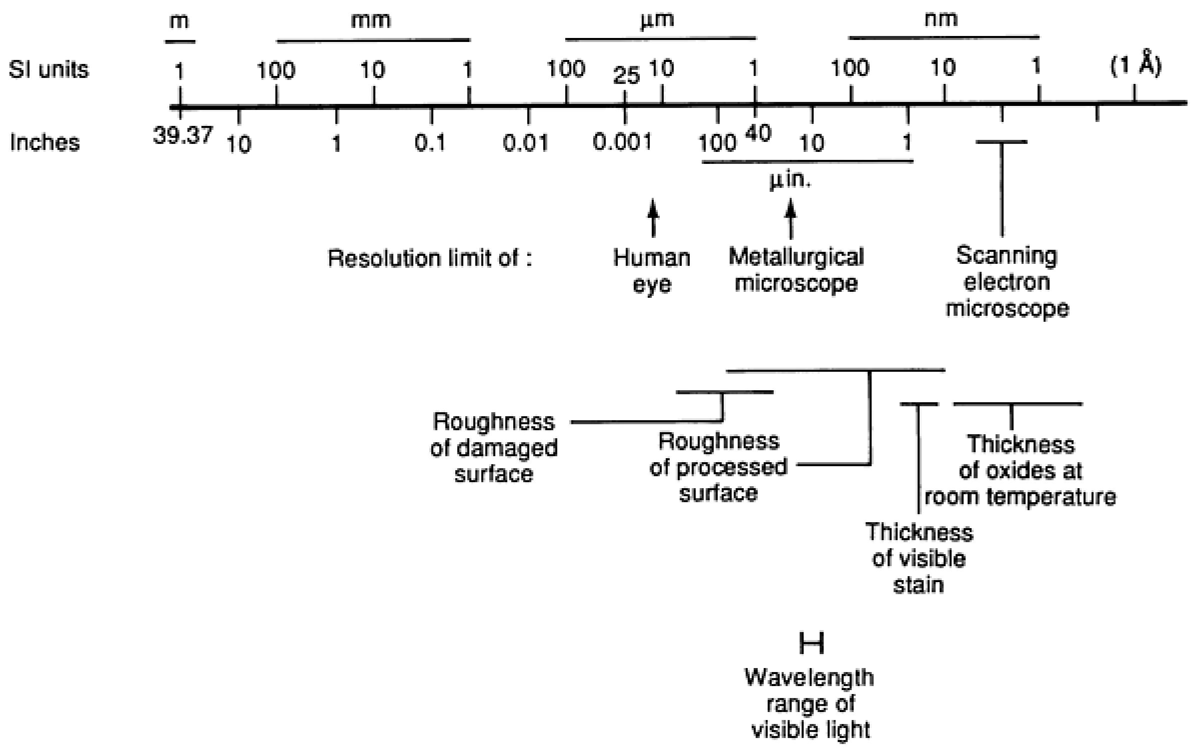


Fig. 1 Resolution ranges of selected surfaces and coatings relative to the viewing capabilities of various optical devices and the human eye itself.

Lateral Resolution Requirements to Discern Specific Features. For observing cracks, defects, inhomogeneities, plastic strains, and the details of surface damage, instruments with an appropriate lateral resolution must be selected. The question of appropriate resolution relates to the "need to know." For example, consider a crack, which nominally can be described as being in the shape of the letter V. At the crack tip, the size or spacing may be as small as atomic radii (~0.3 to 0.5 nm, or 3 to 5 Å), whereas at the other end it may be visible to the naked eye.

Very narrow cracks may constitute a minimal hazard in a structure, and thus may not be worth looking for. However, when a fatigue mode of wear is encountered, even the smallest crack is of interest.

Material defects approach a few atomic diameters and may be no more useful to observe than are crack tips. Material inhomogeneities are of the order of grain sizes (~1 μm, or 40 μin.), and are often a more important find.

The dimension of wear damage is often large relative to the lateral resolution of instruments. The contact diameter between two hard steel balls of 12 mm ($\frac{15}{32}$ in.) diameter pressed together with a load of 450 N (46 kgf) is about 0.52 mm (0.020 in.). The field of view of a high-power (2000× magnification) optical microscope is ~50 μm (~0.002 in.), whereas the field of view of an SEM unit at 20,000× is about 5 μm (200 μin.). What can be seen at such magnifications? The view can be compared with human observation of the landscape. A person may be living within an interesting geological area without realizing it. Assume that such a person is very familiar with a region within 16 km (10 miles) of his home. If that person could scan the region from a height of 32 km (20 miles), he might discern old lake beds and so on, but from an even higher altitude of 160 km (100 miles) he can detect ancient glacier movements. Similarly, observations at high magnification are likely to cause more confusion than enlightenment. It is best to first scan the entire surface and then to focus in on specific regions to obtain an accurate assessment of the surface.

To obtain maximum usage of the equipment used to examine a surface for flaws, a preliminary inspection of the entire surface should be conducted in order to not overlook isolated areas containing defects not specifically being targeted or sought.

Selected References

- D.H. Buckley, *Surface Effects in Adhesion, Friction, Wear and Lubrication,* Elsevier, 1981
- J.I. Goldstein, D.E. Newbury, P. Echlin, D.C. Joy, C. Fiori, and E. Lifshin, *Scanning Electron Microscopy and X-ray Microanalysis,* Plenum Press, 1981

Vibration Analysis

H.R. Martin, University of Waterloo

Introduction

ON-LINE MONITORING of wear conditions between surfaces is an important factor in maintaining efficient plant operation. The progression of wear in bearings, gears, and other surfaces that move relative to each other can lead to a wide range of failure conditions.

It is the interaction between surface asperities, which is due to relative motion, that induces vibrations in the surrounding structure. In turn, this presents a unique distribution, usually spectral, that often directly relates to the wear (damage) in progress. If there is sufficient energy transfer, then sensors placed on the external surfaces of machines can detect vital information regarding the running condition of the machine.

Wear, in a sense, can be categorized into three groups. The first group is undesirable wear that needs to be minimized, such as in bearings and gears. The second group is deliberate wear that results from the manufacturing process, such as grinding. The third group involves processes such as drilling and milling, which result in wear of the tool while metal is being cut away. In the case of the first and third groups, excessive wear will lead to reduced productivity or, in the extreme case, catastrophic failure.

Surfaces in Contact

Machined and even ground surfaces are neither perfectly flat nor smooth. In fact, only 0.1% of the nominal contact area actually touches, under normal loading conditions (Fig. 1). When these surfaces move, the asperity tips alternately weld and then break off. In addition, loading tends to increase the contact area and, hence, the number of tips in contact. Most surfaces in contact are lapped, honed, or ground, and therefore exhibit randomness in the direction of the machining process.

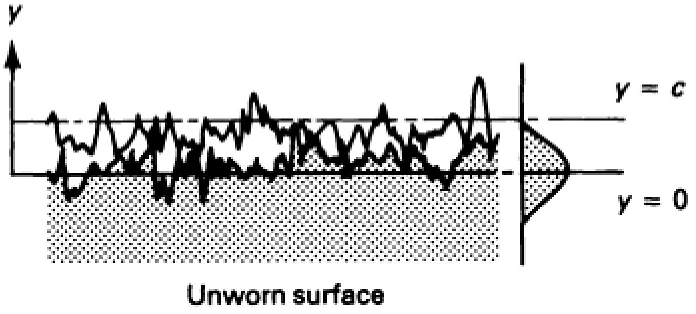


Fig. 1 Typical machine surface finish magnified to show asperity contact

Barwell *et al.* (Ref 1) showed that when the amplitude of the asperity heights was plotted on standard probability paper, a Gaussian distribution results. An example of this (Fig. 2a) shows a ground surface measured on a profiling instrument. For comparison, a surface cut in the form of a sine wave is shown in Fig. 2(b). Hence, it is not surprising that when two such surfaces either roll or slide over each other, vibrations are produced.

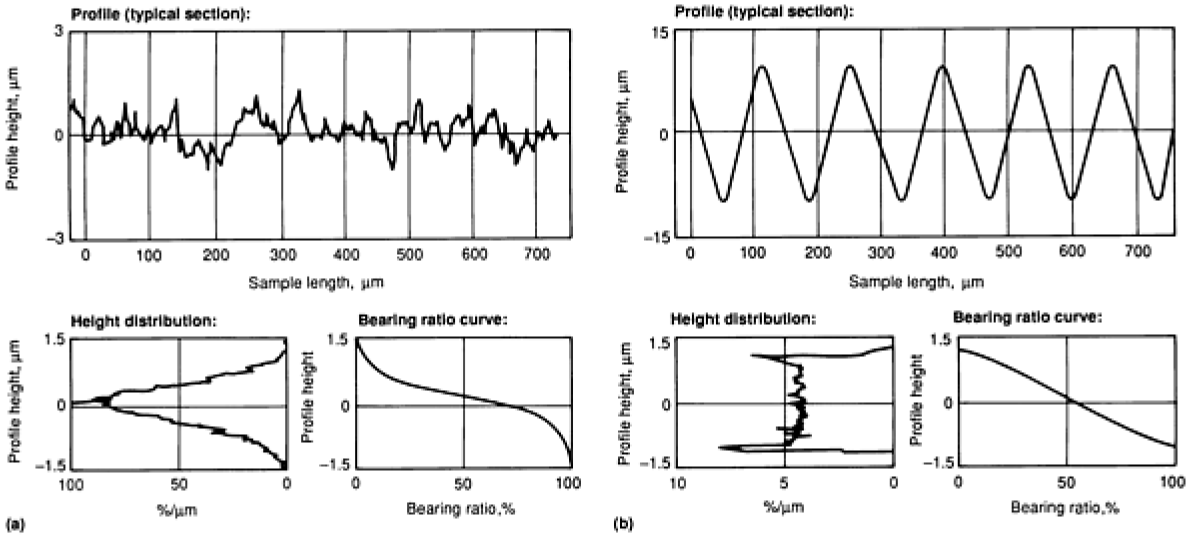


Fig. 2 Examples of statistical distributions for two surfaces. (a) Ground, single direction. Parameters (cut off = 0.8 mm, or 32 mils): R_a = 0.35; R_q = 0.45; R_{max} = 2.31. (b) Sinusoidal cut. Parameters (same cut off as above): R_a = 5.24; R_q = 5.99; R_{max} = 20.6. Source: Ref 2

The other aspect of surfaces in contact involves the deliberate removal of metal, as occurs in a lathe operation. Here, the interest is in monitoring the wear of the cutting tool, because a worn tool will not only reduce productivity, but will also affect the quality of the product. Again, vibrations that relate to the various stages of tool wear, up to breakage, are generated in the structure of the machine.

For example, as a drill becomes dull, it does not cut as fast as it is being pushed into the metal, which results in bending. The resulting "hits" on the side of the hole, which are due to the two lands of the drill bit, produce a spectral signature that is quite different from that of a sharp tool.

On-line too wear sensors are based on the frequency analysis of vibration signals. However, much progress can yet be made in this area (Ref 3). Current approaches still rely heavily on empirical analysis, in conjunction with some kind of database.

In wear monitoring, because the areas of interest in the vibrational spectra are primarily random in nature, the problem of normalizing the features to make them relatively machine-element independent is yet to be solved satisfactorily. Monitoring the trends of such extracted features can probably reduce the need for extensive databases.

Instrumentation

The availability and flexibility of instrumentation for vibration analysis has dramatically improved in recent years with the introduction of microcomputer technology. Transducers used in the measurement of vibration are commercially available. Their principles of operation are based on electromagnetic, electrodynamic, capacitive, resistive, and piezoelectric techniques.

Of these, piezoelectric devices are favored for most general vibration measurement applications, because of their ruggedness, light weight, small size, and ability to self-generate a signal. Additionally, it is possible to design these devices to be resonance-free over a wide frequency range and to be temperatures stable. Other advantages are that both velocity and displacement can be obtained by electronic integration of the signal, although there may be a loss in accuracy.

The accelerometer is an effective vibration transducer over a wide frequency range, but it is not as effective at low frequencies, because the signal magnitude is proportional to the square of the frequency. Below 10 Hz, it is better to use strain gages or other displacement measuring devices.

Another disadvantage of piezoelectric devices is the extremely low power level available, which is due to the nature of the quartz crystal or barium titanate material used. It is therefore essential that a conditioning amplifier with a very high input impedance be used, so that virtually no current is drawn from the sensor.

A popular choice is the charge-type conditioning amplifier, which satisfies this requirement and allows long cable lengths between the transducers and the amplifier, as well, Its disadvantage is poor performance at low frequencies. If, however, this type of performance is required, then the voltage amplifier has to be used with the piezoelectric transducer and the leads kept short.

Unlike the traditional piezoelectric materials, which are hard and brittle, polymeric piezoelectric film (Ref 4), has been available commercially. These polyvinylidene fluoride materials are tough, flexible, lightweight, and very corrosion resistant. The output charge level is about 20 times higher than conventional devices, and the manufacturer claims frequency bandwidths well into the acoustic emission range, that is, above 100 kHz (Ref 5). These developments point to some exciting possibilities for future research.

Modern digital oscilloscopes and analyzers usually have the facility for connection to microcomputers via various serial or parallel communication protocols. This allows great flexibility in data storage, and more extensive analysis.

Analysis Methods and Problems

The raw time-domain record of a random-type signal produced by surfaces in contact that move relative to one another is usually not very informative. Therefore, the traditional approach has been to convert to the frequency domain, either by filtering the signal or by processing using a Fourier transform. This allows the analysis of the vibration spectrum over the

range of frequencies of interest. Usually, the level of energy in a particular frequency band is monitored over time; a significant increase in magnitude is indicative of wear (damage).

The time-domain data collected by a transducer can be transformed faster using the fast Fourier transform (FFT) algorithm (Ref 6). Although this algorithm speeds up the process significantly, especially for large numbers of data points, it is necessary to ensure that anti-aliasing protection is implemented, and that the number of points in a data block is some power to two. However, there is at least one subroutine program that can take a data block with different numbers of points from some power of two.

Because the number of values calculated in the FFT analysis is equal to the number of data points in the original block, they can be paired to form a complex number, so that both amplitude and phase angle can be extracted. Hence, from a block of time-domain data, only half the number of spectral lines are produced in the frequency domain.

In random vibration, such as wear analysis, the phase angle has no meaning. Therefore, some advantage can be gained by using the Hartley transform. Because this algorithm only deals with the real part, some gain in processing speed can be attained. The square of the spectral line amplitudes produces the power spectrum of the data, which is also called the autospectrum. Some relationships that are useful are:

$$\text{Number of spectral lines} = \frac{1}{2} \times \text{number of data points in block}$$

$$\text{Frequency range} < \frac{1}{2}\ \text{sampling rate}$$

$$\frac{\text{Resolution}}{\text{Spectral line}} = \frac{\text{Frequency}}{\text{Number of lines}} = \frac{(2 \times \text{frequency range})}{\text{Number of data points}} = \frac{1}{\text{Block period}}$$

$$\text{Block period} = \text{Number of data points} \times \text{sampling interval}$$

Note that for the number of spectral lines, some commercial analyzers halve the number again, before presenting the video display. For the frequency range, a value of 2.56 is often used.

Most transducers produce analog signals at their outputs, but the computer needs each data point in binary coded form. This requires that the signal be sampled at equal time intervals and digitized. To reproduce the original signal, there are now only a finite number of data points available.

When a continuous signal from a sensor is broken up into a series of discrete numerical amplitude values at discrete intervals, it is called a time series. The process of obtaining the time series from a continuous function is called sampling. In order for the data to be meaningful in the frequency domain, the sampling rate must be greater than twice the highest frequency of interest. The frequency of the signal that corresponds to the minimum sampling rate is known as the Nyquist frequency. The corresponding allowable maximum time between samples is the reciprocal of this frequency.

An analog-to-digital (A/D) converter is needed between the transducer and the computer, where the sampled values are converted into equivalent numerical values by the A/D. Each value is represented by a finite number of states (0 or 1) is a series of steps. These binary coded steps always represent two raised to some integer exponent. Hence, for a 12 bit A/D converter with a full-scale deflection (FSD) of 10 V, the data will have 4096 steps of 0.00244 V. Each sample value is now rounded off to the nearest number of units in the range. This leads to a measurement error that may be as much as one half of the step interval.

The reason for keeping the sampling rate at greater than twice the highest frequency of interest is to avoid aliasing, which is a phenomenon that occurs because false information is generated when the process of interests is sampled (observed) often enough. As shown in Fig. 3(a), when the sampling rate is exactly twice the frequency of interest, no useful information can be extracted. Even if the sample is shifted in time, the correct amplitude of the original analog waveform cannot be fully described. In Fig. 3(b), the sampling rate is too low in relation to the frequency of interest.

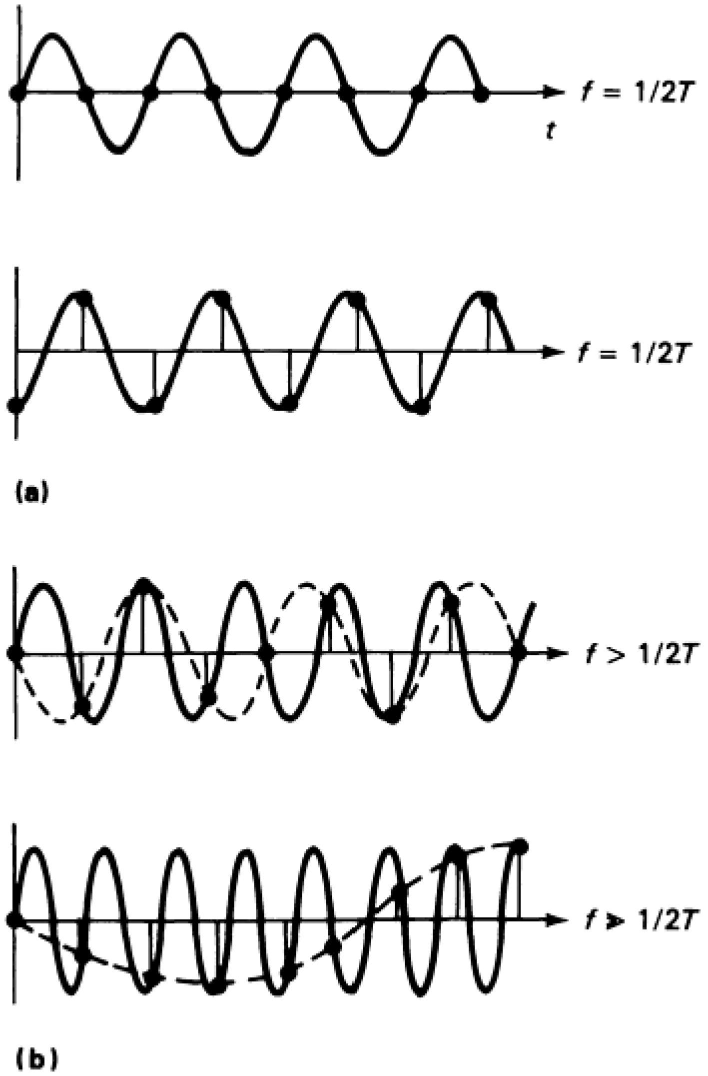


Fig. 3 Examples of aliasing. (a) Extraction of useful information not possible. *f,* frequency; *T,* time period in seconds. (b) Frequency created that does not exist in original data

In each case, the dotted lines show that a second waveform can be drawn through the same points. When processed through a spectral analyzer, this would result in spectral lines that did not exist in the original data. Aliasing is prevented by filtering the signal to a specific bandwidth consistent with the available sampling rate.

Leakage and Windowing

The FFT algorithm operates on the assumption that the collected time record from the sensor is continuous in time, as shown in Fig. 4(a). In the case of a sine wave, there are no problems provided that an integer number of cycles fits into the time-record block exactly. If the input is not periodic within the time block, then the FFT assumes a distorted input, as shown in Fig. 4(b).

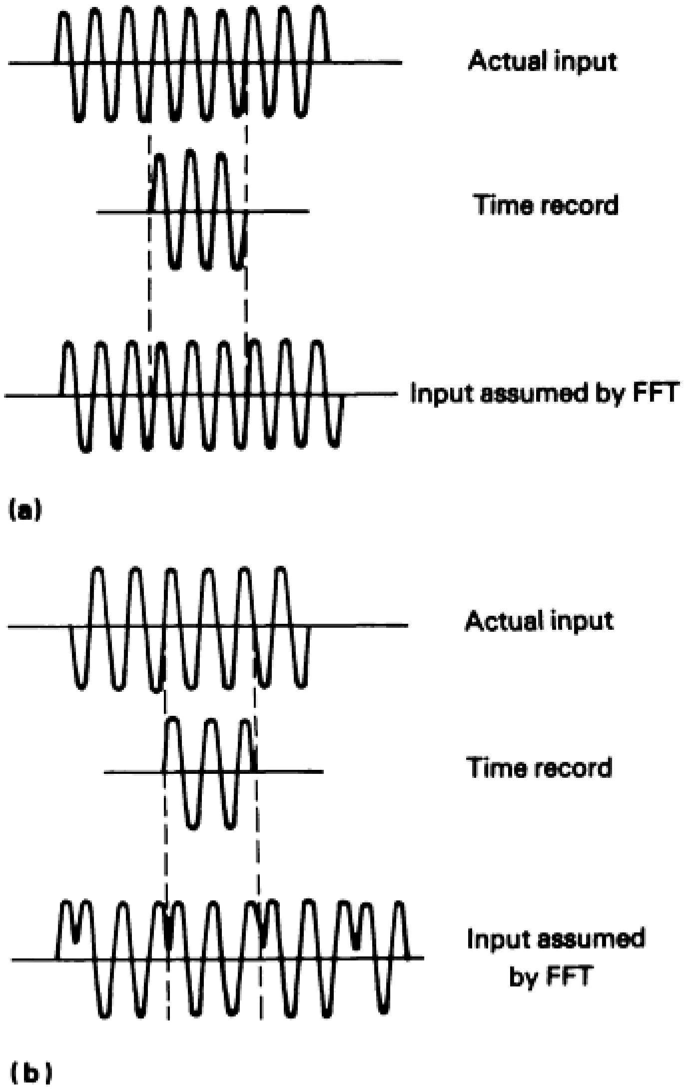


Fig. 4 Examples of signals within time record block. (a) Periodic. (b) Not periodic

In the frequency domain, the spectral energy is not a clean-cut spectral line (Fig. 5a), but is smeared throughout the frequency range (Fig. 5b). This effect is called leakage. It is therefore important to realize that leakage is due to the fact that the data is not periodic within the time block, and can be severe enough to entirely mask small signals close to the main component of interest.

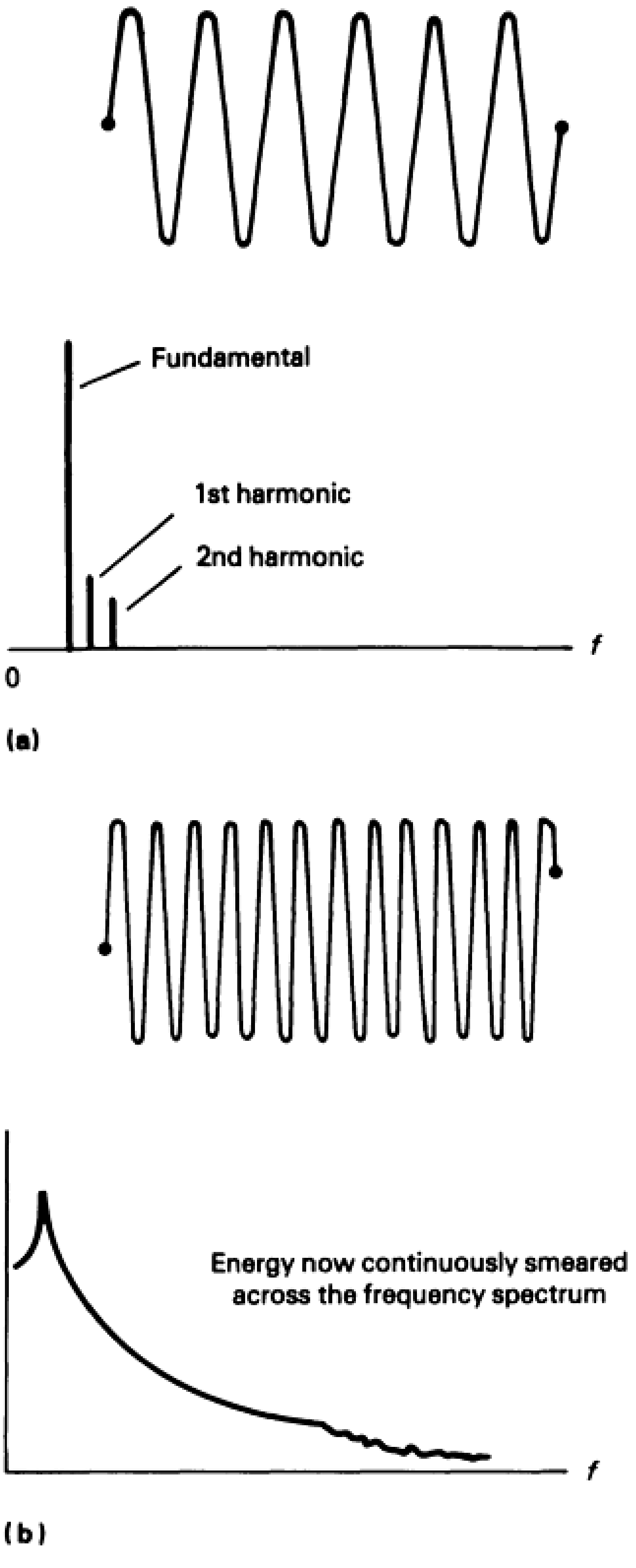


Fig. 5 Analysis results for extreme cases. (a) Data periodic in time block. (b) Data not periodic in time block

The method used to overcome the problem is the use windowing. Examination of Fig. 4(b) shows that most of the problem occurs at the edges of the time block. If the FFT could ignore the edges of the time block and concentrate on the middle portion, then the result should be much closer to the true spectral line. This can be achieved by multiplying the time record by a function that is zero at the ends and unity in the middle. Such functions are called window functions, and a popular one is the Hanning. Its effect on typical data is shown in Fig. 6 and 7. The Hanning window is extensively used for random-type data and therefore is the most useful in analyzing wear data.

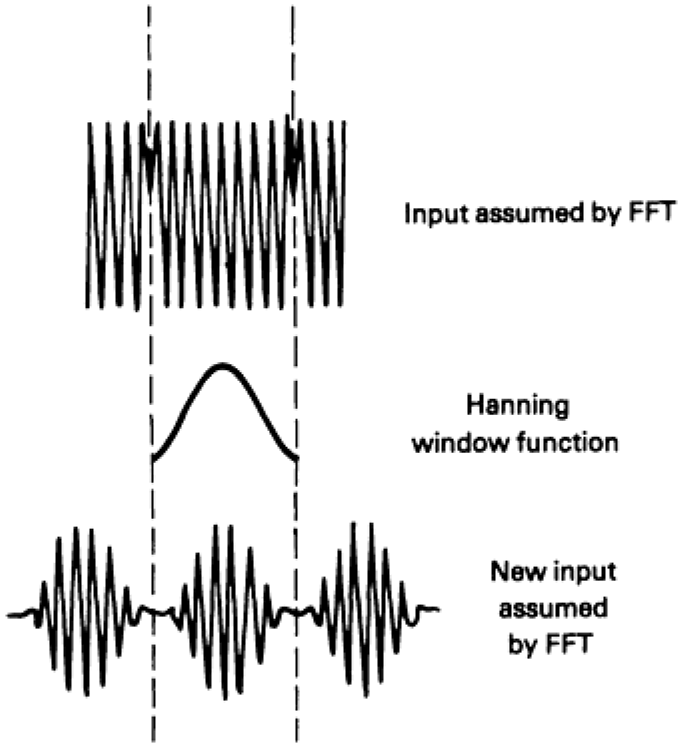


Fig. 6 Effect of windowing on time-domain signal

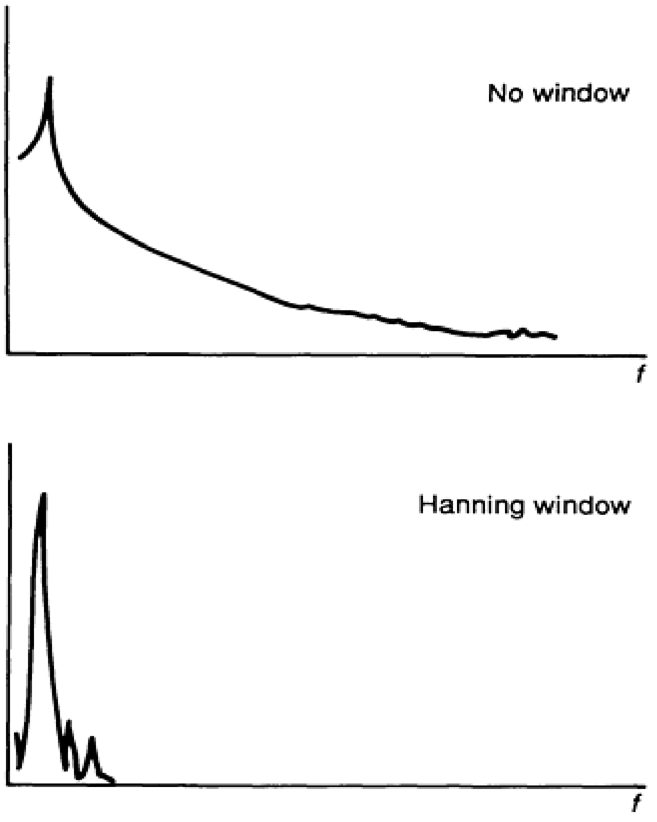


Fig. 7 Effect of windowing on spectrum

Averaging

In many practical situations, data collection has to be carried out in a noisy environment, which means that the signal to be analyzed is contaminated by unwanted signals from other sources. A commonly used method of attenuating unwanted signals is to limit the bandwidth by conventional filtering. Because the noise is usually broadband, this cuts off the extra bandwidth beyond the highest frequency of interest, which eliminates much of the noise. Of course, this approach is of little use if the noise and the signal occupy the same bandwidth. If the unwanted noise signal has a near-zero mean value, then the signal-to-noise ratio (S/N) can be improved by averaging several blocks of data that come from the sensor. Usually, between 4 and 64 data blocks are taken.

Signal averaging is most effective when a deterministic signal is buried in a random signal. It can be applied to the raw time data or to the resultant spectrum of each block. It is a very effective way to clean up (smooth) plots. The improvement in S/N ratio is equal to the square root of the number of blocks of data used.

Random signals cannot be analyzed exactly (unless the sampling time stretches, from minus to plus infinity), unlike deterministic signals, where the sampling time can be confined to one cycle, because of the repeatability of the cycle. Finite sampling times applied to random signalswill introduce errors. If noise is present in the signal, then the quality of the result will depend on the averaging time, T_{av}, and the ideal filter bandwidth, Δf. It has been shown that if the noise in the signal has a Gaussian distribution, then the measurement uncertainty, σ, is (Ref 7):

$$\sigma = \frac{1}{(T_{av}\ \Delta f)^{0.5}} \qquad \textbf{(Eq 1)}$$

For example, if a random vibration is to be analyzed in 20 Hz bands, then the averaging time needed, if the measurement uncertainty is to be less than 5% of the mean squared signal, will be 20 s.

Exponential averaging is a continuous summation average in which most of the recent data are added while earlier data are removed. For example, ten exponential averages means starting with ten summation averages, removing the first, and adding the eleventh, and so on. Therefore, each new piece of data counts for $\frac{1}{10}$th of the average.

Peak averaging considers as many instantaneous spectra as are selected and saves only the highest amplitude at each frequency seen during the data gathering period.

Statistical Analysis Approach

Failure in moving surfaces that are in contact can be due to a number of causes related, primarily, to excessive wear. Pitting occurs at points of maximum Hertzian contact stress. A failure of this type depends on the number of stress cycles, and results in small surface fatigue cracks. Small pits are formed, usually 0.5 to 1.0 mm (20 to 40 mils) in diameter. Large areas of material removal are known as spalling.

Scuffing is related of the lubricating film and is caused by overheating, which could be due to friction and the sliding velocity between surfaces. This type of contact produces alternating welding and tearing, which removes metal rapidly. Once the condition begins, it is very difficult to reestablish a proper oil film. Severe surface welding is called scoring.

Plastic flow is due to cold working of the surfaces caused by high contact stresses, together with rolling and sliding actions. The surface deformation results from the yielding of the surface and subsurface material.

Abrasive wear, unlike the other types of surface damage, is not a local failure, but is likely to spread over large areas.

The amplitude distribution of a vibration signal picked up from surfaces moving relative to each other can be expressed in terms of a probability density function (PDF). This represents an estimate (probability) of the time the signal remains in a particular amplitude window. It has been well established in the literature that machined surfaces are neither perfectly flat nor smooth.

Experiments show (Fig. 2b) that the PDF closely approximates to a Gaussian distribution. As damage (wear) starts to occur, the PDF begins to deviate from this classic bell-shaped curve. Rather than examining the PDF in detail, Dyer and Stewart (Ref 8) decided to track the progress of this damage by examining the statistical moments of the data.

Any signal, particularly a random signal, can be described by its statistical moments. The first three moments are:

- Mean value, or average amplitude size
- Mean squared deviation, or average power in the signal
- Standard deviation, or the measure of how closely the data are clustered around the mean value

A general integral defining all the statistical moments (Ref 9, 10) can be expressed as:

$$M_n = \int_{-\infty}^{\infty} g(y)^n \, p(y) \mathrm{d}y \qquad \textbf{(Eq 2)}$$

where M_n is the nth moment, $n$equals 1, 2, 3, 4, . . ., $g(y)$ is the amplitude function, $p(y)$ is the probability density function, and y is amplitude in millimeters.

For a surface that is in good condition:

$$p(y) = \frac{1}{\sigma \sqrt{2\pi}} \exp \frac{[-(y - \bar{y})^2]}{2\sigma^2} \qquad \textbf{(Eq 3)}$$

Odd moments are related to information about the position of the peak density distribution in relation to the mean value, whereas even moments indicate the characteristics of the spread of the distribution. It has been well established mathematically (Ref 11) that if the fourth moment is normalized using the square of the second moment, then the Kurtosis coefficient takes on the unique value of 3.0, if the surfaces under consideration exhibit a Gaussian distribution of asperities. As the surfaces become damaged, the Gaussian indicator changes value, because the shape of the curve moves from the classic bell shape.

Volker and Martin (Ref 12) showed that for roller bearings, a damage map could be created by plotting the Kurtosis coefficient against acceleration level. It can be seen from Fig. 8 that the types of damage in progress can be identified. The damage map represents a plot of the energy present because of the damage (x axis) versus a measure of the impulsive nature of the damage and the number of defects (y axis).

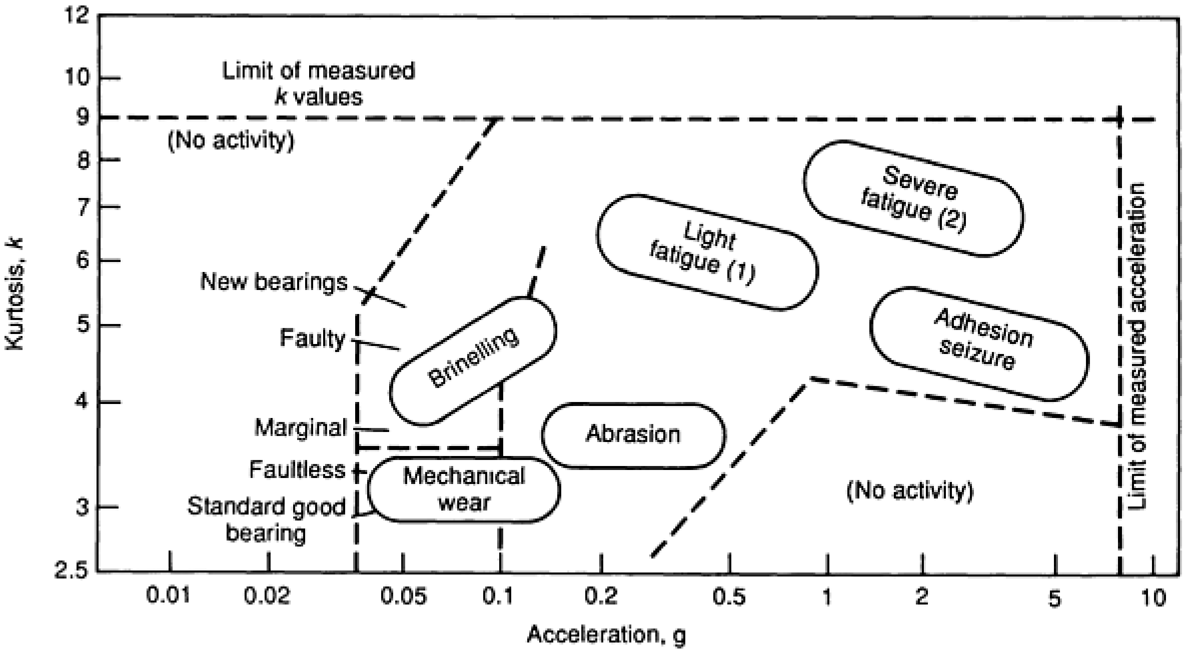


Fig. 8 Damage classification for roller bearings

Figure 9 shows the difference between dry and lubricated bearings using this approach. The accelerometer is placed on the bearing housing and the data is then filtered into five frequency bands on the graph.

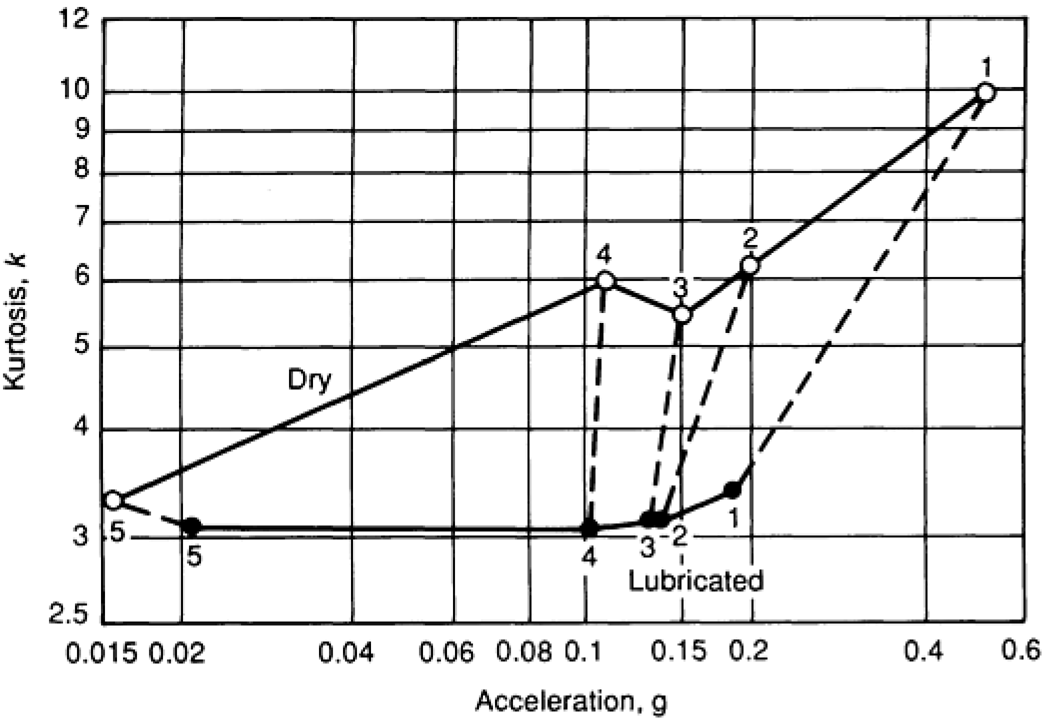


Fig. 9 Difference between dry and lubricating bearing. Points labeled 1 indicate frequency range from 2.5 to 5 kHz; points 2, 5 to 10 kHz; points 3, 10 to 20 kHz; points 4, 20 to 40 kHz; points 5, 40 to 80 kHz

Figure 10 shows the effect of adding silicon powder to the bearing lubrication in order to simulate abrasive damage. The particles used were size #230 SiC, applied in 0.2 mg doses over a period of 8 min.

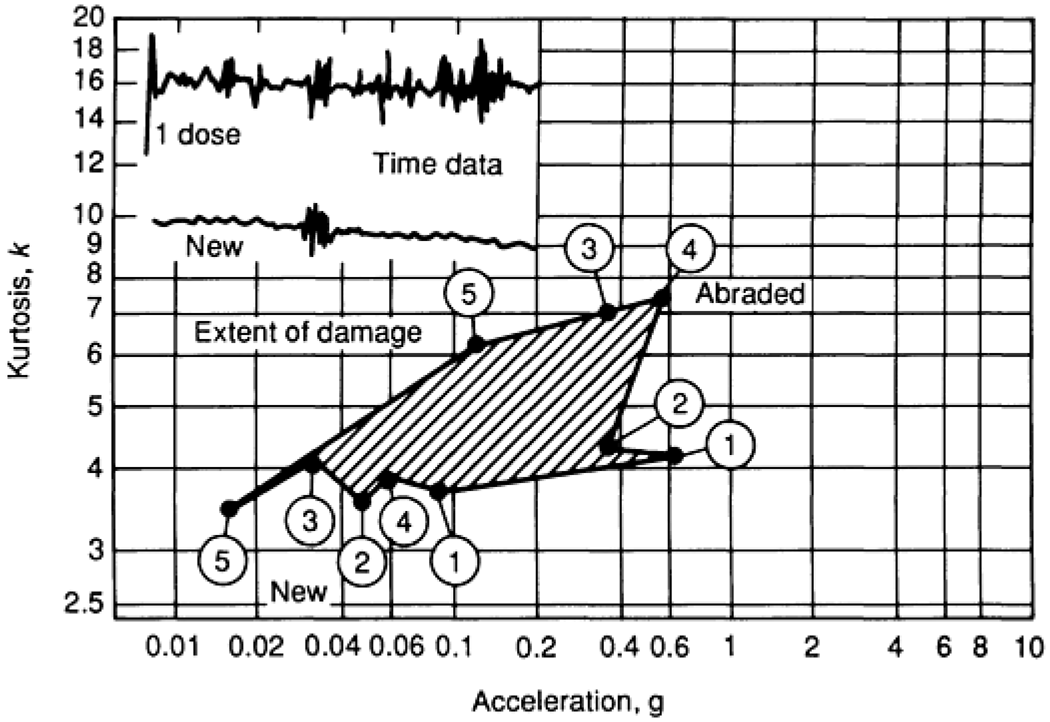


Fig. 10 Effect of adding abrasive powder

Relationship between Friction and Vibration

As mentioned previously, the amount of contact between any two surfaces, in the process of cold welding of the asperities, is a function of the load. In the case of two flat plates, Bowden and Tabor (Ref 13) generated the data shown in Table 1.

Table 1 Amount of contact between two flat plates

Load		True area of contact		Percentage of apparent area of contact[(a)]
N	**kgf**	**cm^2**	**10^{-5} in.^{2}**	
19.6	2	0.0002	3.1	0.0010
49	5	0.0005	7.8	0.0025
980	100	0.01	155	0.05
4900	500	0.05	775	0.25

(a) Nominal contact area, 20 cm^2 (3 in.^{2})

The general relationship between load and contact area can be expressed as:

$$A = \frac{F}{H} \quad \textbf{(Eq 4)}$$

where A is the true area of contact (cm^2), F is the applied load (N), and H is the indentation hardness (N/cm^2). The volume of material worn away when one of the plates slides over the other is proportional to the true area of contact, the total distance of sliding, and the nature of the materials.

$$V = kAl = \frac{kFl}{H} \quad \textbf{(Eq 5)}$$

where V is the volume of material removed (cm^3), l is the total distance traveled (cm), and k is the wear coefficient.

It has been found, in practice, that the wear coefficient, k, remains constant until the pressure between the plates exceeds a value greater than one-third of the hardness, H. Hence, the wear rate is reasonably linear. At pressures in excess of this, k begins to increase, and the wear rate rises rapidly and nonlinearly.

It is suggested that running parts, which are in motion relative to each other, need to be carried out with relatively light loads, until an indication of constant wear rate shows that the surfaces have settled in to each other. This happens when the asperities have been ground down and the effective pressure falls, allowing higher pressures to be applied. A typical value for k would be 26×10^4.

This "running in" effect illustrated in Fig. 11, which shows the results for a new bearing, just out of the box, and a repeat set of results after 60 minutes running. Although the vibration level does not give any significantly information, the Kurtosis values indicate that after 1 h, the results are closer to 3.0. This indicates a reduction in peakness in the data, that is, the asperity distribution is closer to Gaussian, after the honing period.

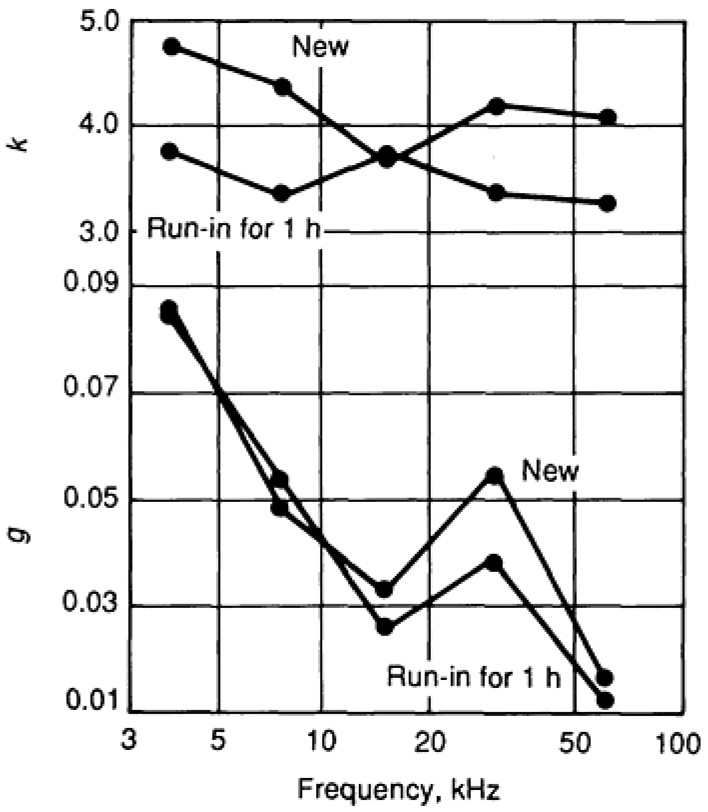


Fig. 11 Detecting the progress of "running in"

The interaction of surface asperities during sliding causes complex vibration patterns in both the normal and tangential directions. It has been suggested (Ref 14) that the asperities act as a system of microsprings, with appropriate stiffnesses in two directions. These normal and tangential contact springs are, of course, nonlinear with loading, because the geometry of the peaks changes, and more of them come into contact. For dry friction conditions, the asperities can be modeled as a random distribution of microaxially loaded bars that are set in free vibration by the sliding motion between the plates.

Calculations based on the above theory and experimental data indicate that the frequency of free vibration for metallic pairs experiencing asymmetrical deformation of asperities was found to be approximately in the range from 500 to 3000 Hz. The higher the relative velocity between the two surfaces, the more prominent is the normal component of impulses, because of collisions between the asperities, and the higher the amplitude of contact vibrations.

References

1. F.T. Barwell *et al.,* The Interaction and Lubrication of Rough Surfaces, *Proceedings of the Symposium of the International Union of Theoretical and Applied Mechanics,* IUTAM (Enschede, Holland), 1974, p 304-329
2. M. Brock, Fourier Analysis of Surface Roughness, *Bruel and Kjaer Technical Review,* No. 3, 1983, p 3-45
3. S. Braun, Ed., *Mechanical Signature Analysis,* Academic Press, 1986, p 321-342
4. J. Chatigny, Piezo Film Yields Novel Transducers, *Electron. Week,* Aug 1984
5. R.V. Williams, *Acoustic Emission,* Adam Hilger, Bristol, 1980
6. J.W. Cooley and J. W. Tukey, An Algorithm for the Machine Calculation of Complex Fourier Series, *Maths of Computation,* Vol 19 (No. 90), 1965, p 297-301
7. J. Bendat and A. Piersol, Chapter 6, *Random Data: Analysis and Measurement Procedures,* Wiley-Interscience, 1971
8. D. Dyer and R. Stewart, Detection of Rolling Element Bearing Damage by Statistical Vibration Analysis, *J. Mech Design (ASME),* Vol 100, 1978
9. J.S. Bendat, *Principles and Applications of Random Noise Theory,* John Wiley & Sons, 1958
10. A. Papoulis, *Probability, Random Variables and Stochastic Processes,* McGraw-Hill, 1965
11. H.R. Martin, Review of Gear Damage Monitoring Techniques, *Proceedings of the First International Machinery Monitoring and Diagnostic Conference* (Los Angeles), Society for Experimental Mechanics, 1989, p 183-189
12. E. Volker and H.R. Martin, Application of Kurtosis to Damage Mapping, *Proceedings of the Fourth International Modal Analysis Conference* (Los Angeles), Society for Experimental Mechanics, 1986, p 629-633
13. F.P. Bowden and D. Tabor, "Friction and Lubrication," Methuen Co., 1956
14. B.V. Budanov, Mutual Relation between Friction and Vibration, *EuroTrib 81,* Vol 1A, 1981, p 240-246

Selected References

- K.G. Beauchamp, *Signal Processing,* Allen and Unwin, 1973
- E.O. Brigham, *The Fast Fourier Transform,* Prentice-Hall, 1974
- I.P. Castro, *An Introduction to the Digital Analysis of Stationary Signals,* Adam Hilger/ESM, 1989
- N.H. Cook, Tool Wear Sensors, *Wear,* Vol 62, 1980, p 49-57
- G. Kivenson, *Durability and Reliability in Engineering Design,* Pitman, 1971
- W. Lenkiewicz, The Sliding Friction Process--Effect of External Vibration, *Wear,* Vol 13 (No. 2), 1969, p 99-108
- P.A. Lynn, *Electronic Signals and Systems,* MacMillan, 1986
- J.S. Mitchell, *Machinery Analysis and Monitoring,* PennWell Books, 1981
- D.E. Newland, *An Introduction to Random Vibrations and Spectral Analysis,* 2nd ed., Wiley, 1984
- A. Papoulis, *Probability, Random Variables and Stochastic Processes,* McGraw-Hill, 1965
- R.T. Spurr, Frictional Oscillations, *Nature,* Vol 169, 1961
- T. Vinh and J. Blouet, Non Stationary Signal Processing--Applications to the Study of Time Dependent Sliding Friction, *Annals CIRP,* Vol 30 (No. 1), 1981

Lubricant Analysis

F.E. Lockwood, Pennzoil Products Company; R. Dalley, Predict Technologies

Introduction

LUBRICANT ANALYSIS PROGRAMS for predictive maintenance are now routinely practiced by a number of industries, including railway, aircraft, automotive and truck, chemical, refinery, and various others (Ref 1 2, 3, 4, 5, 6, 7, 8, 9, 10, 11). The physical appearance, viscosity, metals content, and other properties of oils sampled periodically from operating equipment are interpreted to provide an early warning of impending failure or to signal the need for oil changes and routine maintenance. Best known of such programs are the spectrometric oil analysis program (SOAP) and related programs practiced by the U.S. military services (Ref 12). Similar programs have also been developed for commercial aircraft, commercial truck and auto fleets, and industrial equipment such as compressors, gearboxes, and pumps. Continuous on-line oil monitoring may one day be standard practice.

Oil/wear particle analysis also is often a valuable failure analysis tool, although less has been published on this subject (Ref 13, 14). Failure analysis generally is practiced by mechanical engineers who primarily examine hardware. However, in the area of automotive lubricants, extensive use of oil/wear particle analysis is made in "post-mortem" failure investigations. This expertise is similarly useful in many other areas.

These two subjects, covering the details and references of applications of oil/wear particle analysis, are the focus of this article. The important oil analysis methods will be reviewed, and appropriate test programs for predictive maintenance and a methodology for failure analysis will be discussed.

Acknowledgements

Oil analysis cases and a number of helpful comments were provided by C.M. Comer, J.W. Fu, T.E. Rushing, and J. Torres of Pennzoil Products Company. Extensive information and figures were extracted from the *Wear Particle Atlas,* currently a publication of Predict Technologies.

Oil / Wear Particle Analysis Methods

Oil/wear particle analysis to determine lubricant condition includes both physical inspection and chemical examination of wear debris, contaminants, and reaction products from lubricants such as engine oils, hydraulic fluids, cutting fluids, greases, and synovial fluids from humans and animals (Ref 15). Because wear is an inevitable and anticipated consequence of surface contact between interacting machine parts such as shafts, bearings, gears, and bushings even in properly lubricated systems, oil/wear particle analysis can potentially be applied to all lubricated equipment. Equipment life expectancies, safety factors, performance ratings, and maintenance recommendations are predicated on normally occurring wear and lubricant service. Such factors as design complexity, unit size, intricacy of assembly configurations, and variations in operating conditions and environments can make maintenance or repair needs difficult to evaluate or detect without taking equipment out of service. Oil/wear particle analysis allows noninterruptive diagnostic determination of lubricant condition by determining the amount of wear and the lubricant reaction products. Based on this determination, equipment condition or impending failure can be predicted.

Sampling of Service Lubricants. Lubricant analysis begins with the sampling procedure, and the validity of a particular analysis depends on how well this procedure is carried out. The ideal sample is taken immediately downstream from the lubricated surfaces--for example, from a drain line off an individual bearing, prior to filtration, while the equipment is operating under usual conditions and temperatures. Care is taken to obtain a representative sample by discarding any volume that may have been stagnant in the drain line. The sample is captured in a clean nonmetallic container, sealed, and carefully labeled, including information about lubricant and equipment history. In practice, it is difficult to achieve this ideal. Samples often must be taken from sumps and recycle lines or large reservoirs. Deficiencies

in sampling points, however, are to some extent compensated for by consistent sampling at the same point under the same operating conditions (Ref 5).

Once in the laboratory, all samples should be brought to a uniform temperature and stirred condition before testing. This is particularly important when studying lubricant additive and wear metals, which may stratify in the lubricant under some conditions.

Sampling frequency is another key concern. This depends on the type of equipment, service conditions, and critical nature of service. Equipment maintenance records should suggest a proper sampling frequency. Otherwise, it is a good idea to sample frequently (for example, weekly) until a track record is built; then, if desirable, the rate of sampling can be lessened. Once a possible problem is detected, the sampling frequency must be increased until a positive determination is made on equipment condition and the action to be taken.

As general guidelines, aircraft should be inspected after every flight, while large industrial equipment with good filtration systems may require only monthly inspection. One rule of thumb is 25 to 40 h sampling frequency for gas turbines and 100 to 500 h for diesel engines. Recommended inspection guidelines for checking the condition of lubricants used in large industrial equipment are given in Table 1.

Table 1 Recommended lubricant inspection intervals for selected engines, drive systems, and power generating units

System	Recommended operating intervals between inspections, h
Aircraft gas turbine	50
Airborne hydraulic system	50
Aircraft derivative gas turbines	50
Diesel engine	200
Heavy transmission/gears	200
Surface hydraulic system	200
Heavy-duty gas turbine	200-500
Steam turbine	250-500
Large-bore reciprocating engine	250-500

Source: Ref 6

For each lubricant parameter that is measured, a control record is built that after a period of time will reveal normal operating ranges for a given type of equipment/lubricant. For parameters such as viscosity, direct comparison with lubricant and equipment manufacturer specifications also provides information on the acceptable operating range. In setting up a sampling and analysis program, it should be kept in mind that unless parameter operating guidelines are known beforehand, the program must provide enough information to set statistical guidelines for acceptable versus abnormal parameter limits.

Physical Inspection. A simple physical inspection, particularly in the case of failure analysis, can speed diagnosis of poor equipment operation and wear. Typical lubricant indicators that are often observed are listed in Table 2.

Table 2 Diagnostic guidelines for detecting and rectifying service lubricant deterioration

Lubricant appearance	Instrument verification of lubricant deterioration		Corrective action
	Test objective	**Detection method**	
Visibly thin or less viscous than fresh lubricant	Check for fuel dilution	Gas chromatography	Replace or vacuum strip lubricant contaminated with fuel
Visibly thick or more viscous than fresh lubricant; oxidized odor is detected and lubricant color is much darker than fresh lubricant (severe oxidation is evidence of excessive drain periods, abnormally hot running conditions, or exposure to abnormal types and levels of preoxidation)	Confirm lubricant oxidation level	Infrared spectroscopy	Replace lubricant, eliminate factors that accelerate oxidation
Milky lubricant (typically indicates formation of water emulsion)	Confirm presence of water	Infrared spectroscopy, Karl Fischer titration, hot-plate sputter test	Replace or vacuum strip lubricant, eliminate source of water if possible
Unusual precipitates or gel structures present in lubricant (due to contamination or presence of other lubricants)	Identify contaminants after filtration	Infrared spectroscopy	Replace lubricant, identify and eliminate source of contamination, evaluate and change lubricant formulation if necessary

Physical Testing. The most common physical tests run in conjunction with spectrometric and wear metal analysis programs are viscosity, total acid number (TAN), and determination of water content. ASTM methods are normally used for measuring viscosity (ASTM D 445) (Ref 16), TAN (ASTM D 974 or D 664) (Ref 17, 18), and low concentrations of water (Karl Fischer Titration) (ASTM D 1744) (Ref 19). The presence of glycol coolants can also be detected via ASTM method D 2982-85 (Ref 20). In cases where water is present at levels above 0.05 vol%, infrared spectroscopy is usually used. Although control limits for deviation of each of these parameters need to be set depending on the type of lubricant and equipment, viscosity variations of ±20%, TAN of greater than 3 mg/g, and water (in oil-based lubricants) in excess of 100 to 500 ppm are normally cause for action or at least further investigation. The presence of coolant is always a cause for concern, because it implies that coolant is leaking into the lubricant.

Spectrometric Metals Analysis. In the late 1940s, the railroad industry began testing lubricants for wear metals. With the advent of practical emission spectrometers, the SOAP methods were developed for military aircraft and then extended to gasoline- and diesel-powered military vehicles. Spectrometric methods include atomic absorption (AA), atomic emission spectroscopy (AES), inductively coupled plasma emission (ICPE) (Ref 21, 22), and x-ray fluorescence (XRF) (Ref 23). Of these methods, AES and ICPE, which rely on the detection of light emitted by the elements, are the most popular because of cost, speed, and other factors.

Unless special research measures are taken (Ref 24, 25, 26), spectrometric metals analysis determines the concentration of soluble metals and metal particles up to 10 μm in size. Therefore, it follows mild (benign sliding) rubbing wear and the early stages of fatigue quite well, because in these wear modes the predominant distribution of wear particles is within the detectable (<10 μm) range. However, in abnormal wear situations, such as severe sliding, rolling fatigue, cutting and abrasive wear, and scuffing wear, particles are generally larger than 10 μm. In such situations, which are typically of greater interest, ferrography (wear particle analysis) and particle counting are useful and, in some respects, superior monitoring techniques (Ref 3). Ferrography operates in the most useful particle size range, determining relative concentrations of midsize (<1 to 250 μm) particles; particles larger than 100 μm can be determined by particle analyzers. Combinations of these techniques are claimed to be the most effective (Ref 2), but spectrometric analysis has been most popular because of its relatively low cost. A further advantage is that a number of contract laboratories offer this type of analysis with very rapid turn-around time. In the near future, the particle size range of spectrometry may indeed be extended, making it more competitive with ferrography (Ref 27).

Spectrometric metals analysis has been successfully used to monitor aircraft engines. One correlation of maintenance findings to spectrometric analysis found a 70% prediction rate of malfunctions (Ref 28). On the other hand, the cost effectiveness of the technique is questionable, because overhauls may sometimes be undertaken sooner than necessary.

The application of spectrometry to gasoline and diesel engines has been somewhat more difficult because of greater complexity of contamination sources from fuel dilution and blow-by products from combustion gases, and the greater range of engine operating parameters. Test programs invariably include other physical and chemical tests to determine

viscosity change, water and fuel dilution, coolant contamination, and chemical changes in the engine oil. In general, such programs have been applied more successfully to rail and commercial diesel fleets or stationary diesels (Ref 29) than to spark-ignition engines, and the cost is more easily justified. Diesel engine bearing failures in military ground equipment have been predicted (Ref 12). Both abnormal bearing wear and piston/liner scuffing have been detected in advance in locomotive diesels (Ref 1) (see the section "Applications of Ferrography" in this article).

Iron, copper, lead, chromium, and aluminum are the principal component metals analyzed. Tin, silver, nickel, molybdenum, titanium, and vanadium are usually of lesser significance but are sometimes present, generally as alloying elements or coatings. The presence of silicon or the combination of silicon, aluminum, and titanium is often sought as evidence of dirt contamination. Contaminants from engine coolants include boron, potassium, and sodium. Lubricant additive elements analyzed are zinc, calcium, sodium, copper, magnesium, chlorine, phosphorus, antimony, molybdenum, sulfur, and boron. Because some of these elements fit into more than one category (for example, copper may be a wear metal and a lubricant additive metal), comparison with the baseline unused lubricant is desirable. Comparison allows easy identification of wear trends as well as changes in additive concentration between the used and new lubricant.

Table 3 summarizes several guidelines for recognizing normal versus abnormal operation. It must be emphasized, however, that trends should be established for the particular equipment under study, because material composition, operating conditions, lubricant, sump size, and absence or presence of filtration all determine "normal" levels of metals.

Table 3 Selected lubricant indicators and range of sensitivities required for indicator detection

Indicator	Indicates	Investigative level
Elements		
Iron	Wear	100% increase or change of 10-20 ppm[a]
Copper	Wear	100% increase
	Additive	100-200 ppm typical[b]
Aluminum	Wear	100% increase
	Additive	As specified
	Dirt	>25 ppm[c]
Chromium	Wear	100% increase
Molybdenum	Wear	100% increase
	Additive	100-200 ppm[d]
Lead	Wear	100% increase
	Fuel	100% increase
Tin	Wear	100% increase
Silver	Wear	100% increase
Nickel	Wear	100% increase
Vanadium	Wear	100% increase
Titanium	Wear	100% increase
	Dirt	>25 ppm
Silicon	Dirt	>25 ppm
	Additive	As specified
Sodium	Coolant	>30 ppm
	Additive	1000 ppm[d]
Boron	Coolant	>30 ppm
	Additive	1000 ppm[d]
Magnesium	Additive	500-1500 ppm[d]
Calcium	Additive	500-2000 ppm[d]
Zinc	Additive	1000 ppm[d]
Phosphorus	Additive	1000 ppm[d]
Barium	Additive	As specified
Antimony	Additive	As specified
Potassium	Coolant	>30 ppm
Sulfur	Additive	5000 ppm[d]
Chlorine	Contaminant	>20 ppm
	Additive	<200 ppm
Nitrogen	Additive	500-1500 ppm
	Fuel oxidant	2× normal
Compounds		
Water	Contamination	>100-500 ppm
Physical properties		

Viscosity	Oxidation or fuel oil	Change of ±20%
Total acid number (TAN)	Oxidation	>3 mg/g
Testing procedures		
IR ferrography	Wear[(e)]	±20% change in absorbance
	Dirt[(e)]	Increase of 50-100% in D_L (large particles)
Particle size analysis	Wear	100% increase in particle count or increase in number of large particles

(a) Wear metals should be investigated for cause when a sharp (>100%) increase occurs from baseline or a change of 10-20 ppm or more occurs.

(b) Typical levels are given for passenger car motor oils. Additive levels departing ±30% from norm should be investigated.

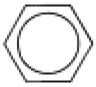
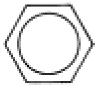
(c) Change from additive baseline is considered.

(d) Typical values.

(e) See Table 4 in this article.

Infrared (IR) spectroscopy is widely used to determine water and coolant contamination of the lubricant, as well as to identify (Ref 30) and monitor the depletion of additives and the buildup of oxidation products. The availability of Fourier transform infrared spectroscopy (FTIR) allows the detection of small changes in the IR spectrum of the used lubricant. A differential spectrum can be obtained by subtracting the spectrum of the new lubricant from that of the used lubricant to clearly reveal the areas of change. Table 4 indicates the IR absorbance wave numbers typically monitored.

Table 4 Infrared spectroscopy of lubricants

Contaminant or lubricant component	Adsorption, cm^{-1}	Type of vibration	Interferences
Water	3300-3500, 1600	O−H stretch	Glycol, oxidants
Fuel	500-1000, 1600	C−H out of plane deformations of aromatics	Unsaturation in lubricants
Oxidized lubricant	1700-1750, 1100-1200	C=O stretch	Viscosity index improvers
	1630, 1270	N=O stretch (oxidized from blow-by)	Alkenes
Glycol coolant	1040, 3300-3500	C=O stretch	Oxidation
	1080, 3300-3500	C−O stretch	Antioxidant
ZDDP	950-1050	P−O−C asymmetric stretch	Aromatics
Sulfonate detergent	1100-1200	SO_3 asymmetric	Oxidation
Pour-point depressant (polymethacrylate)	1732		
		C=O stretch	Oxidation
Dispersant:		O O	
Succinimide	1707	‖ ‖	
Mannich base	1650	C−N−C	Oxidation
Viscosity index improver:			
Polymethacrylate	1732	 −O−R	
Styrene-isoprene	700	C=O	
		 −R	Oxidation, Fuel

Contaminants such as water and glycol can be picked up in infrared by detection of the -OH stretch for water and the -C-O stretch for glycol. Degradation due to oxidation is detected as the carbonyl, -C=O, stretch from the formation of organic acids and conjugated carbonyls. Lubricant oxidation will also result in the depletion of antioxidant additives in the lubricant, usually zinc dialkyldithiophosphate (ZDDP) or phenolic compounds. The ZDDP concentration can be monitored via the P-O-C stretch; however, the phenolic antioxidants are usually not present in great enough concentrations to be easily differentiated from other compoundsadsorbing at the same wavelength. Several other additives can usually be detected, including detergents, dispersants, and occasionally polymethacrylate pour-point depressants, depending on concentration.

Infrared equipment manufacturers are quite familiar with lubricant analysis and can provide assistance in setting up analysis methods. Some IR manufacturers have designed equipment with oil analysis especially in mind. In addition, lubricant analysis by infrared is widely available at contract laboratories.

Particle counting, which involves monitoring the number of particles of a given size range per fluid volume, has been used as a primary monitoring tool in combination with other analytical methods. Both particle counting and direct-reading ferrography detect the onset of severe wear as a rapid increase in the amount and size of particles. Particle counting detects all particles, whereas ferrography screens primarily for ferrous wear particles.

Typical reporting formats require that the number of particles per milliliter of fluid volume be broken down into the following particle size categories based on the method used (Ref 31):

Particle size, μm	
ISO method	**OSU method**
>5	>10
>10	>20
>15	>30
>25	>40
>50	. . .

Particle-detection equipment is either of the light-interruption type or the laser-scanning type. Problems with this method include the difficulty of obtaining consistent samples and incorrect counts caused by artifacts such as air bubbles. A more laborious, but acceptable, alternative to particle counting is collecting the particles onto a filter (sediment testing) and then counting and sizing them under a microscope (Ref 11, 31, 32).

Particle counting is widely used to monitor hydraulic systems, where wear particles larger than 10 μm (beyond normal spectrometric limits) are of primary interest (Ref 11). Many manufactures publish recommended particle-count levels (Ref 31). Particle counting is more suitable when fatigue mechanisms are a primary means of failure or when contaminant particles cause abrasion. Fatigue of subsurface origin occurs in full-fluid-film conditions, leading to larger initial debris distribution, which may proceed rapidly. This scenario, where λ(the ratio of oil film thickness to surface roughness) exceeds approximately 1, is amendable to monitoring by particle count or magnetic plug detection (Ref 33).

Magnetic plug/chip detection (MCD) is a variation of the filter/counting method of particle counting. A magnetic stub is introduced into the oil flow in a piece of machinery to continuously collect ferrous material. The debris typically is viewed microscopically. This collection method favors large debris and is therefore suitable for systems that run in the full-fluid-film condition (for example, many hydraulic systems).

Ferrography. First adopted by the U.S. military, ferrography is now also used by heavy-equipment manufacturers, oil companies (Ref 5, 34), and other industries for wear particle analysis and predictive maintenance. The technology, developed by Westcott and Seifert (Ref 35, 36, 37) in the early 1970s, is used either as the primary analytical method or in conjunction with spectrometric analysis (Ref 2) and has been shown to give earlier prediction (Ref 5) and greater diagnostic information than spectrometric analysis alone. However, ferrography is comparatively expensive--on the order of $25 per routine sample if performed in house versus less than $10 per sample for spectrometric analysis (1990 prices). It is not particularly useful when applied to equipment where the samples must be taken from extremely large sumps (large steam and power generation systems, for example).

Ferrography is unique in that it allows potential determination of the amount and type of wear as well as the source of wear. Ongoing advances in foregoing instrumentation have enabled the broader study and classification of wear particles produced by many different metals and substances, both magnetic and nonmagnetic.

A ferrographic analysis of wear particles begins with the magnetic separation of machine wear debris from the lubricating or hydraulic media in which it is suspended. To establish accurate baselines for oil condition, regular samples are taken from carefully selected locations with the machine system, preferably during normal operation, as described in the discussion of sampling earlier in this article. In ferrographic examination, two types of ferrographs may be used. The direct-reading (DR) ferrograph uses optical density to quantitatively measure the concentration of wear particles in a

lubricating oil or hydraulic fluid. The particles are subjected to a powerful magnetic-gradient field and are separated in descending order of size (Fig. 1). Particles 5 μm and larger are confined to the entry end of the deposition field. The particle sizes become progressively smaller along the deposition path. Light attenuation at two locations along the path--at the entry deposit, D_L (large particles), and a point several millimeters farther down the tube, D_S (smaller particles)--is used to quantify the relative concentration of "direct large" (D_L) to "direct small" (D_S) particles. Values of wear particle concentration (WPC) and the percent of large particles (PLP) are thereby derived, establishing machine wear baselines and trends in wear condition.

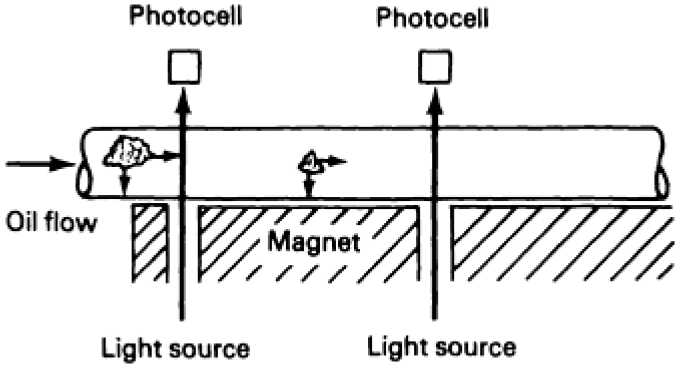


Fig. 1 Diagram of DR ferrograph deposit areas

Machines starting service go through a wearing-in process, during which the quantity of large particles quickly increases and then settles to an equilibrium concentration (Ref 38) during normal running conditions. A key concept of ferrography is that machines wearing in an abnormal mode will produce unusually large amounts of particles and a particle distribution with proportionally more large particles; that is, both WPC and PLP will show a significant increase above the baseline established after wearing-in (Fig. 2). The different regimes of wear, from mild to severe, a characterized by different size particles, the most severe being associated with particles larger than 1 mm (Ref 39).

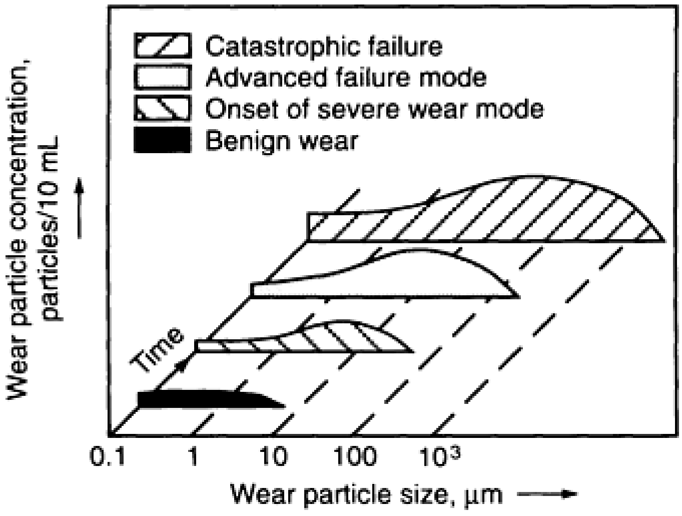


Fig. 2 Typical progression of severe wear

When subsequent DR ferrographs indicate an abnormal trend in wear, analytical ferrographic techniques can be utilized to study the wear pattern to specifically identify the nature of potential machine problems. In analytical ferrography, ferrograms (slides upon which wear particles have been deposited) are prepared and examined under a bichromatic microscope for measurement and identification.

Preparation of the sample entails (1) dilution of used fluid (lubricant, hydraulic fluid, or aqueous solution) with a wash solution to improve particle precipitation and adhesion and (2) flow of the prepared sample down and inclined slide, passing across the magnetic field. Wear particles arrange themselves along the slide, with the largest particles deposited first (Fig. 3). Ferrous particles line up on strings that follow the magnetic-field lines of the instrument. Nonferrous particles and contaminants travel downfield in a random distribution pattern not orientated by the magnetic field. This long deposition pattern spreads the wear particles out, providing good resolution of large and small particles--an important factor in diagnosing wear problems.

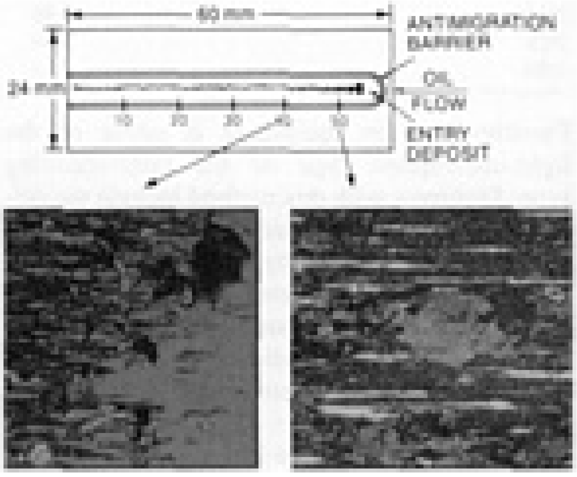


Fig. 3 Typical analytical ferrogram deposit patterns

The resulting ferrograph is examined using a three-power bichromatic microscope with cameras. Under magnification of 100×, 500×, and 800×, the microscope utilizes both transmitted and reflected light sources together with red, green, and polarizing filters to distinguish the size, composition, shape, and texture of both metallic and nonmetallic wear particles. The wear particles are classified to determine the type of wear and its source.

Types of Wear Particles

Ferrography is used to differentiate abnormal wear conditions from the normal rubbing wear that occurs during stable machine operation and from break-in wear that occurs during start-up of equipment. During break-in of a wear surface, the rougher surface irregularities are smoothed by grinding contact of the two surfaces so that the surfaces conform or "mate," effectively forming a smoother contact region. During this time, some large particles will break off the surfaces; however, in metals the majority of the particles formed result from exfoliation of the ductile, mechanically worked layer that is created at the metal surface. The slow process of forming and rubbing off of this shear mixed layer results in many small, normal rubbing-wear particles. When this layer is removed too rapidly (by abrasion, for example) or when other undesirable processes take place (fatigue, for example), new types or unusual amounts of wear particles are detected by ferrography.

The various types of wear particles are illustrated in Fig. 4. These particles are described in considerable detail in Ref 39, 40, and 41.

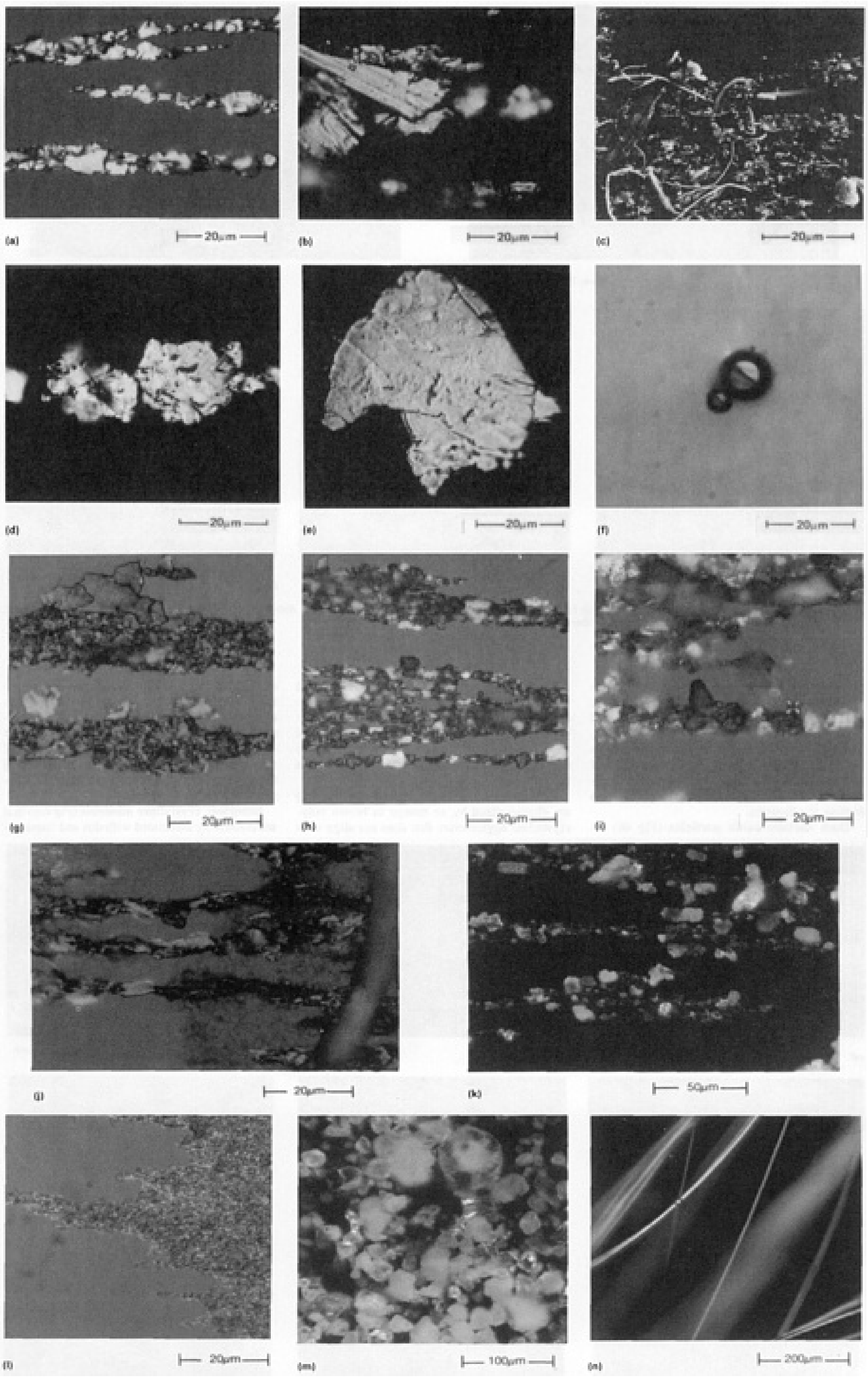


Fig. 4 Various types of wear particles. (a) Normal rubbing wear particles. (b) Sliding wear particles. (c) Cutting wear particles. (d) Fatigue particles. (e) Laminar particles. (f) Spheres. (g) Red oxide sliding wear particles. (h)

Dark metallo-oxide particles. (i) Black oxide particles. (a-i) 1000 ×. (j) Friction polymer. 1000×. (k) Red iron oxide particles. 400×. (l) Corrosive wear particles. 1000×. (m) Inorganic crystalline debris (road dust). 200×. (n) Glass fibers. 100×

Rubbing wear particles consist of flat platelets (Fig. 4a), generally 5 μm or smaller, although they may range up to 15 μm before their wear effect is considered to be severe. There should be little or no visible texturing of the surface, and the thickness should be 1 μm or less. A special case of normal rubbing wear is break-in wear, which is characterized by long, flat particles generated as machining marks are rubbed off by sliding surfaces. Abrasive contaminants can also dramatically increase the amount of rubbing wear, occasionally to the point of causing failure.

Sliding wear particles (Fig. 4b) are identified by parallel striations on their surfaces. They are generally larger than 15 μm, with the length-to-thickness ratio falling between 5 and 30. Severe sliding wear particles sometimes show evidence of temper colors, which may change the appearance of the particle after heat treatment.

Cutting wear particles (Fig. 4c) may resemble wire, drill turnings, whittling chips, or gouged-out curls. They may be caused by penetration of a soft surface by a hard, sharp edge, perhaps resulting from fracture of another component, or by cutting by abrasive particles embedded in an opposing soft surface.

Fatigue spalls and chunks (Fig. 4d) are removed from the metal surface as a pit or crack opens up. They are generally larger than 5 μm, with a length-to-thickness ratio of less than 5. There is generally some surface texture, and particles appear rough and shaped like chunks of coal, rather than flat. Spalls are similar in appearance to chunks, but are thinner. Small spalls are distinguished from normal rubbing wear by slightly greater thickness and surface texture. It is often necessary to examine very small particles at magnification of 800× to clarify these characteristics.

Laminar Particles. When a particle of any severe wear type passes between the surfaces of rolling elements, the effect is similar to that of a rolling pin on pie dough. The particle is flattened out, the edges may split, and there are often holes in the center. Particles such as this are called laminar particles (Fig. 4e). The length-to-thickness ratio is generally greater than 30. Although laminar particles can be very small, in a practical sense only the larger particles will be rolled out. Laminar particles larger than 15 to 20 μm are indicative of the formation of other severe wear particles. The presence of laminar particles in addition to spheres is indicative of rolling-bearing fatigue microcracks.

Spheres (Fig. 4f) are caused by wear, fatigue, or contamination. Their formation as a wear phenomenon is generally associated with rolling elements. Spheres formed by wear mechanisms are generally less than 5 μm in diameter, with very smooth surfaces. Such spheres are often a precursor of fatigue spalling. If the diameter exceeds 5 μm, or if the surfaces appear rough or oxidized, the spheres were probably caused by cavitation or contamination. Source of contamination include grinding and welding. Spheres in this size range can also be formed during lubricant oxidation, but can be distinguished by identifying their chemical composition, which will typically involve metallic lubricant additives.

Red oxide sliding wear particles (Fig. 4g) resemble severe sliding wear particles, except that they are usually gray in color and, when viewed in white transmitted light only, appear translucent and reddish brown. They are formed in conditions of inadequate lubrication and are, in effect, severe sliding wear particles that have oxidized, the oxide being Fe_2O_3. Particles of this type that are thick and rounded (with a thickness ratio similar to chunks) may originate from fretting mechanisms.

Dark metallo-oxide particles (Fig. 4h) resemble red oxide sliding wear particles, except that they contain a core of free metal and thus are not translucent. They also often show flecks of free metal on their surfaces. These particles are caused by heat and lubricant starvation, and indicate more severe wear. Large, partially oxidized particles indicate catastrophic surface failure.

Black oxide particles (Fig. 4i) are dark gray to black in color and resemble pebbles in shape. The oxide in this case is Fe_3O_4. Such particles result from a more severe condition than red oxide particles, in that a proportionally greater amount of iron is consumed in the oxidation process because of inadequate lubrication.

Friction polymer (Fig. 4j) is a material that forms when a lubricant is under stress; the resulting polymeric material is insoluble in the solvents used in ferrography. Depending on the wear mechanisms within the equipment, there may or may not be metal particles trapped within the polymer.

Red iron oxide (Fe_2O_3) particles (Fig. 4k) are characterized by an orange to brown polycrystalline agglomerate that does not align with the magnetic field on the ferrogram. They typically result from the presence of water in the oil. Their color may best be evaluated under reflected polarized light. Particles that change from yellowish orange to a more reddish brown after heat treatment are hydrated iron oxide and probably originate from rust. Particles that are reddish brown before heat treating may be rust that has been exposed to a drying mechanism such as heat, or they may originate from fretting or other corrosion/oxidation mechanisms.

Corrosive wear particles (Fig. 4l) result from attack on the surfaces of a machine and its wear particles by acids and other corrosive agents. They are submicron-size particles of free metal, oxides, and other metal compounds and are so small that they generally do not form a deposit along the ferrogram. However, in the eddy currents at the exit from the ferrogram, and under the influence of the magnetic flux at the end of the magnetic field, a deposit of this material will form. The size of this deposit can warm of chemical attack on the equipment.

Inorganic crystalline minerals (Fig. 4m) that are commonly associated with dirt and construction materials will depolarize light that has passed through a polarizer. This phenomenon is called birefringence. Materials that are birefringent usually show some degree of internal order, or crystallinity. The birefringence of inorganic materials usually is not influenced by heating to the temperatures used for analyzing a ferrogram. Some minerals are not birefringent and thus must be classified under the "Other particles" category below.

Organic crystalline materials that are birefringent include certain plastics, wood, Teflon, insect parts, and cotton. These materials will generally char or lose birefringence upon heating to 345 °C (650 °F).

Other particles that may appear in ferrogram as contamination include glass, amorphous blobs, paper dust, paint, varnish, glue, and so on. Other particles that may appear as part of a lubricated system that do not fall into any of the specific classification given here include molybdenum disulfide, graphite, and seal materials.

Fibers (Fig. 4n) include any particles that are fibrous, even if they fit into categories such as organic crystalline. Typical fibers include hair, cotton (from rags), wood, glass, minerals (rock wool), nylon (from brushes, and cellulose (from filters).

Alloy Identification: Heat Treatment of Slides. For the purposes of alloy identification ferrography, direct chemical analysis by techniques such as AES or ESCA (electron spectroscopy for chemical analysis) can be used, but are rarely available. Heat treatment of the particles on the ferrogram slide is a quick, inexpensive method of identification. Heating the particles at 340 °C (640 °F) for 90 s yields oxide film thicknesses that are in the range of the wavelengths of visible light. Reflection of light off the metal surface underlying the oxide layer produces interference effects, resulting in coloring of the particles. Different classes of alloys exhibit predictable colors. Therefore, the prior heat history of a particle may sometimes be apparent as temper colors or variations in the color of the heat-treated surface. Table 5 provides a guide to alloy identification based on heat treatment at 340 °C (640 °F) for 90 s.

Table 5 Application of ferrography to identify ferrous and nonferrous alloy wear particles

Heat treated at 340 °C (640 °F) for 90 s

Wear particle		Identifying color
Material	**Composition or property**	
Low-alloy steels	<1% C	Turns blue under heat treatment; degree of color saturation varies, depending on residual oil on the particles and variations between alloys
Medium-alloy steels	Generally cast iron or case-hardened low-alloy steel with <3.5% C, without other alloying elements	Turns a straw color under heat treatment
High-alloy steels [a]	. . .	Under heat treatment, no significant change in color is observed. Weakly affected by the magnetic field of the ferrograph, and thus shows a more random distribution across the ferrogram than other ferrous alloys
Copper alloys	. . .	Show few variations under heat treatment. Brass generally shows no change from its characteristic yellow color, or a slight deepening of the color toward gold. Aluminum bronze may show a mottled appearance with varying intensities of yellow, white, and blue.

		Copper alloys are not affected by magnetic fields and will be randomly distributed on the ferrogram.
Aluminum	. . .	Alloys do not show color change under heat treatment. Aluminum can be confused with high-alloy steel; less dense than steel, particles will tend to appear farther down the ferrogram for a given size distribution than steel. Has a more whitish cast than does stainless steel. Other chemical tests can be performed.
Tin/lead alloys	. . .	Dull or grayish cast before heat treatment. Heating metals the alloy and facilitates oxidation of the entire particle, most often resulting in a whitish residue at the site of the original particle.
Other alloys	. . .	See table of tests for white nonferrous alloys in Ref 42.
Molybdenum disulfide	Commonly used as an EP additive in greases and may be found in other lubricants. Diamagnetic, which causes MoS_2 particles to collect near the nonwetting barrier or, if large enough, to distribute evenly over the slide	Particles are bluish gray, are unaffected by heat treatment, and have a layered, somewhat crystalline, appearance.

(a) Includes all stainless steels

Applications of Ferrography

This section examines specific concerns regarding wear particles and examples of the application of ferrography to various types of equipment.

Gear Boxes. Monitoring of gear boxes and detection of gear wear within motors (Ref 3) and other industrial equipment constitute a broad area of ferrography application. In gear wear, combined rolling and sliding occur, leading to variations in characteristic particles, depending on relative rolling and sliding velocities. A gear pair fails only when teeth break, preventing transmission of power, or when noise and vibration become unacceptable. Therefore, ferrographic observation of gear wear must be placed in the context of the entire affected operating system when deciding upon the need for immediate maintenance.

The most common problem observed is fatigue at the pitch line (Fig. 5a and b), where rolling occurs. Pits form in the pitch-line area either initially (arrested pitting) or continuously, resulting in eventual tooth damage. The formation of a large number of pits may result in a step at the pitch line. Another type of fatigue process is the exfoliation of the skin on hardened gears, which occurs on nitrided, carburized, and other hardened materials.

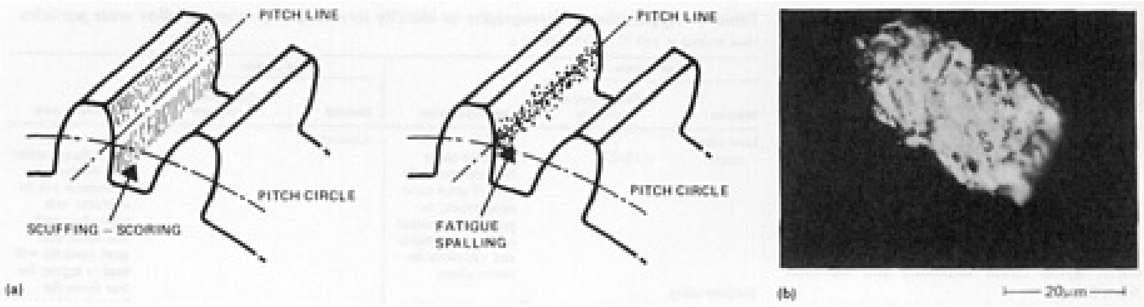


Fig. 5 (a) Gear failure. (b) Chunky fatigue particle from gear pitch line. 1000×

In the sliding portion of the gear (Fig. 5a), scuffing or scoring may occur, resulting in sliding wear particles with striated surfaces. These conditions, detectable by ferrography, indicate the need for correction, often by use of a more viscous lubricant or one with a more active extreme-pressure (EP) additive in order to prevent severe scuffing and total failure. Other modes of gear wear include overload wear and abrasion by contaminants.

To handle the problem of elastohydrodynamic (EHD) lubrication effects, the designer needs to consider three regimes of lubrication (Ref 43):

- *Regime I:* no appreciable EHD oil film (boundary)

- *Regime II:* partial EHD oil film (mixed)
- *Regime III:* full EHD oil film (full film)

Figure 6 shows idealized gear system operating regimes as a function of speed and load. To the left of the overload wear curve, where heavy loads are carried at low speed, wear occurs because the EHD oil film becomes discontinuous. At higher speeds, the allowable load increases because the EHD oil film is partial (due to partial metal-to-metal contact) or full (thick enough to prevent metal-to-metal contact). Above the fatigue spalling line, wear is governed by the strength of the gear material. It is not that the lubricant is inadequate, but rather that the load is transmitted through the oil film. If the load is excessive, fatigue particles from the gear pitch line will be generated. If the load is greater still, a tooth may break. Choice of lubricant has little effect in this case, because this event is governed primarily by material selection and load. If speed is increased, the wear regime will be to the right of the scoring or scuffing line.

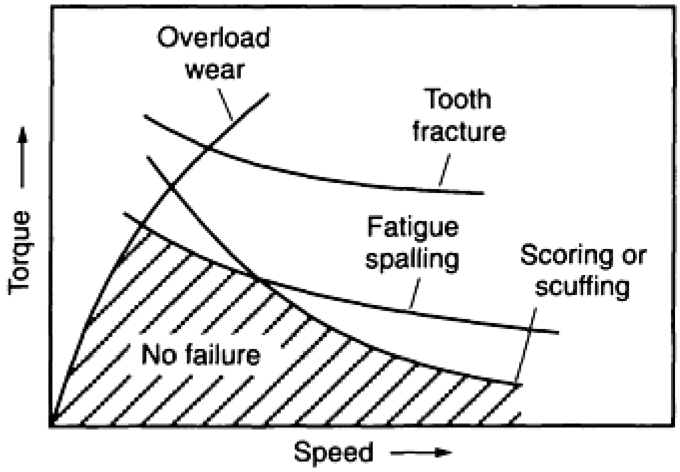


Fig. 6 Gear system operating regimes

Fatigue particles from a gear pitch line have much in common with rolling-element bearing fatigue particles. They generally have a smooth surface and are often irregularly shaped. A high ratio of large particle to small particles is found by direct-reading ferrography, which is again similar to rolling-element bearing fatigue (Fig. 4d). Depending on the gear design, the particles may have a major dimension-to-thickness ratio between 4:1 and 10:1. Chunkier particles (Fig. 5b) result from tensile stresses on the gear surface, causing the fatigue cracks to propagate deeper into the gear tooth prior to spalling.

Scuffing of gears is caused by a load and/or speed that is too high. Excessive heat generation breaks down the lubricant film and causes adhesion of the mating gear teeth. Roughening of the wear surfaces ensues, with subsequent increase in wear rate. The regions of the gear teeth affected are between the pitch line and both the gear root and tip.

Once initiated, scuffing usually affects each tooth on a gear, resulting in a large volume of wear debris. Since there is a large variation in both sliding and rolling velocities at the wear contacts, corresponding variations in the characteristic of the particles are generated. The ratio of large to small particles in a scuffing situation is low. All the particles tend to have a rough surface and a jagged circumference. Even the small particles may be distinguished from rubbing wear by these characteristics. Some of the large particles have striations on their surfaces, indicating a sliding contact. Because of the thermal nature of scuffing, quantities of oxide are usually present, and some of the particles may show evidence of partial oxidation--that is, tan or blue temper colors. The degree of oxidation depends on the lubricant and the severity of scuffing.

The ratio of large to small particles depends on how far the surface stress limit is exceeded. The higher the stress level, the higher the ratio becomes. If the stress level rises slowly, a significant increase in the quantity of rubbing wear prior to the development of any large severe wear particles may be noticeable.

Severe sliding wear particles are 15 μm in size or larger. Some of these particles have surface striations as a result of sliding. They frequently have straight edges, and their major dimension-to-thickness ratio is approximately 10:1. As the wear becomes more severe, the striations and straight edges on particles become more prominent and many large particles showing striation marks are present.

Diesel Engines. Ferrography for diesel engines has been used in conjunction with other test methods (usually spectrometric analysis) with superior results (Ref 44). As in other oil-lubricated equipment, wear is indicated by increasing amounts of particles and by changes in particle size distribution, composition, and morphology. The effects of engine operating conditions on the wear of cylinder liners, piston rings, and crankshaft main bearings have been successfully observed via ferrography (Ref 44).

For diesel engines, heat treatment of analytical ferrographs distinguishes between, for example, low-alloy steel (crankshafts) and cast iron (piston rings and cylinder liners), depending, of course, on the specific engine metallurgy. Although ferrous particles are primarily analyzed, other particles such as lead may be partially retained and have been used to follow main bearing wear (Ref 44).

Normal ferrograms from diesel engines generally show only small rubbing wear particles and very few large metal particles. A light deposit of corrosive wear particles at the ferrogram exit is typical. Diesels are exposed to acid conditions caused primarily by sulfur-containing fuels. In the United States this is becoming less of a problem, because the sulfur content of diesel fuel is being reduced by environmental regulation. Common wear problems in diesel are bore polishing, in which the cylinder wall is polished in spots to a mirror finish, and ring wear. Both of these problems are associated with piston deposits to some degree, depending on the engine. This wear mechanism results in an increase in wear debris that is detectable by both ferrography and spectrometry, since the number of small particles increases. Another problem--cylinder wall scuffing or scoring--tends to occur in engines with tight crownland clearances. This problem can also be detected by ferrography as the presence of an increased number of large particles with striations, typical in sliding wear. Similarly, ferrography has been used to detect scuffing in a diesel valve train system where sliding was taking place (Ref 45).

Aircraft Gas Turbine Engines. Aircraft and aircraft-derivative jet engines are subject to various failure mechanisms. Some of these failure modes proceed very rapidly, whereas others can be detected hundreds of operating hours before a shutdown condition is reached. Most failures of gas turbines occur in the gas path. Gas-path failures frequently, but not always, cause an increase in wear particle size and concentration in the oil system, probably due to the transmittal of imbalance forces to turbine bearings and other oil-wetted parts. The resulting bearing or gear wear is then detected by ferrographic observation.

Determining the exact source of a wear problem can be difficult in gas turbines because of the complexity of the oil-wetted path. Typically, several cavities, housing, or gears will be force lubricated through individual return lines connected to a tank from which the oil is pumped, then passed through a filter and heat exchanger, and the cycle is repeated. Magnetic chip detectors or magnetic plugs are often installed in the return lines from the various engine parts. These can help to pinpoint the source of generation in cases where particle metallurgy, as determined by heat-treating ferrograms, is similar for various engine parts. However, chip detectors will not give a warning until the wear situation is so severe that extremely large particles are being generated. By this time, the opportunity for predictive maintenance may be lost. Other analytical techniques, such as spectometric oil analysis or vibration analysis, may help to pinpoint the part(s) in distress; however, in most cases the engine will have to be inspected by borescope or, more likely, by disassembly. In this instance, the oil analysis must be confident that a problem exists.

Nevertheless, the colds of determining a failure before it is critical improve if ferrography and spectrometric analysis are combined as screening tools and backed up by analytical ferrography or microscopic examination of wear debris (Ref 3). Continued analysis of existing gas turbine monitoring programs and improvements in such programs are needed to ultimately determine their cost effectiveness.

Gasoline Engines. In general, ferrography seems to be used far less for gasoline engines than for diesels. However, Fiat has adopted ferrography for the development of new engine components and prototypes, thereby significantly reducing test time and avoiding catastrophic failures (Ref 9). During engine testing, the oil was monitored by ferrography and spectrometry every 20 h. Ring and liner wear was detected, as well as the effect of such test variables as blow-by. Ferrography was found to be far superior to spectrometry in this application.

Hydraulic Systems. Problems in hydraulic systems are most often caused by contamination (Ref 11) or fatigue wear. Because most of the contaminant and fatigue wear particles are larger than 10 μm, the analysis program should begin with a particle-monitoring test--either particle counting. DR ferrography, or sediment testing. In addition, most programs include viscosity, water, and spectrometric metals analysis.

The particularly deleterious nature of particles in the 10 to 40 μm size range has been pointed out by Leugner (Ref 32). Particles of this size are not detected spectrometrically, yet they often escape filtration and circulate with the lubricant, acting as abrasives. Upon determination that a system contaminant problem exists, improved filtration is often part of the solution (Ref 5, 32).

Ferrography has been used extensively by the Fluid Power Research Center to measure the quantity and to identify the type of wear debris in hydraulic system components as well as in complete systems. This work has resulted in an improved hydraulic pump contaminant sensitivity test.

Among other findings, the work demonstrated that for linear mechanisms more wear occurs for dust with a maximum particle size near the spool clearance dimension than for much coarser or much finer dust at the same concentration. This is attributed to the wedging action of critical-size particles that are close to the clearance dimensions. The Fluid Power Research Center also performed ferrographic analysis over substantial portions of the operating life for the following systems: (1) two complete vehicle hydraulic systems (agricultural tractors), (2) a vehicle hydraulic steering system, (3) a vehicle transmission lubrication system, and (4) an auxiliary hydraulic system on an agricultural tractor. Ferrogram readings as well as particle analysis data are presented in Ref 46.

Compressors. Gas compressors of the reciprocal, rotary, or centrifugal type are often candidates for lubricant monitoring programs in industrial plants. Depending on the type of compressor, cylinder wear, rolling-element bearing water, or gearbox wear may be detected by ferrographic analysis. Lubricant viscosity, contamination by water, and chemical composition should also be monitored in many cases, because deleterious effects caused by contact with the gas being compressed are not unusual.

Grease. In order to apply the techniques of ferrography to grease-lubricated bearings, a solvent system must be used that dissolves the grease sample to produce a fluid of suitable viscosity for ferrogram preparation. The ingredients used in grease formulations are diverse, including a wide variety of soaps or thickeners and solid particles such as molybdenum disulfide. However, a mixture of toluol/hexane--an aromatic; aliphatic, essentially nonpolar blend of solvents--has been found to be a good general solvent. Sampling greases is difficult. Typically, only the grease in the immediate vicinity of the wear contact is being worked and thus contains the wear particles to be observed. This grease can normally be sampled only by complete teardown of the bearing. Nevertheless, such analysis can be useful, particularly for failure analysis. Infrared spectroscopy and physical observation often also provide evidence as to the mechanism of grease deterioration.

Case Histories

In practice, equipment is often complex and different types of wear particles are found together.

Gearbox Wear. In a case where severe sliding and overload occurred because of lubricant deficiency in a process industry reduction gearbox, examination of the ferrogram at low magnification showed an obvious abnormal wear mode ongoing because of the many large metal particles present (Fig. 7a).

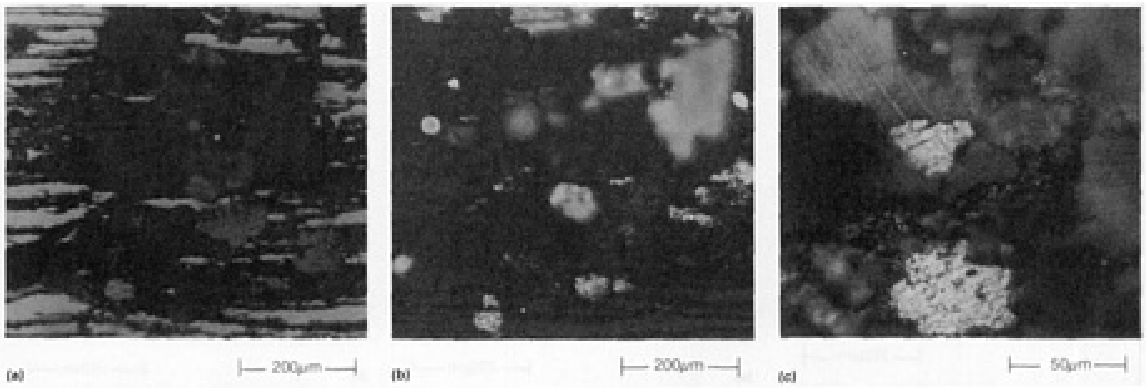


Fig. 7 Gearbox wear. (a) Wear particles resulting from gear overload. 100×. (b) Red oxide particles. Polarized reflected light. 100×. (c) Temper-colored particles. 400×

In this sample, however, may of the severe wear particles showed striation marks, indicating that sliding was involved during their generation. Also, the relatively huge size of some of these particles was indicative of a severe sliding regime. Present, although not as plentiful as the sliding wear particles, were large free metal platelets with smooth surfaces and irregular edges, characteristics typical of rolling-element bearing fatigue or gear-touch pitch-line fatigue.

Therefore, the wear particles appeared to be generated at the pitch as well as at the tips and roots of the gear teeth. Present to a lesser extent were large cutting wear particles and some copper alloy particles. Figure 7(b) shows the entry deposit in polarized reflected light, which emphasizes the presence of large, flat, red oxide agglomerates that may be described as scale; the significance of these particles is relation to the wear situation was not ascertained. Examination of the ferrogram at a magnification of 1000× shows that the free metal wear particles are virtually free of oxidation, such as temper coloring; it can be safely assumed that the abnormal wear was not due to speed-induced scuffing or scoring, as represented on the right side of Fig. 6.

Actual inspection of the same gearbox revealed that the gear teeth were heavily worn--especially at the tips, where the case had been worn away. This fact explains the presence of both steel and cast iron particles in this sample. In the manufacture of gears for industrial applications, a steel gear often is case hardened by heating in a carbon atmosphere so that carbon will diffuse into the outer layers of the gear. Subsequent quenching and tempering of the gear harden the outer case, but leave the steel core soft. This results in a hard, wear-resistant surface with a tough, shock-resistant core to prevent tooth breakage. Examination of a heat-treated ferrogram from such a gear shows the particles to range from blue to straw color, depending on their carbon content. In Fig. 7(c), both straw and blue temper-colored particles are present. The low-alloy steel particles (blue, shown as darker particles) are consistent with the finding that the case had worn away at the tips of the teeth, exposing softer steel. The steel particles show striation marks, indicating a sliding contact.

The problem was solved by using a gearbox oil with EP additive, which arrested the excessive wear. EP additives moved the overload wear curve in Fig. 6 to the left so that, for this case, an operating regime formerly, outside the no-failure envelope was now within it.

Water in the Oil in a Reduction Gearbox. Figure 8 shows the entry in polarized reflected light of a ferrogram prepared from an oil sample from a reduction gearbox used to drive an agitator in a pharmaceutical manufacturing plant. The agitator motor and gearbox were roof mounted, with the impeller driveshaft extending down from the ceiling to the mixing tank inside the plant. In this case, water had entered the gearbox, which was splash lubricated, causing an abnormally high wear rate. Many red oxides, characteristic of water attack, were present. Practically no free metal wear particles were found in this sample, probably the result of oxidative attack caused by the water during the two-week storage time prior to preparation of the ferrogram.

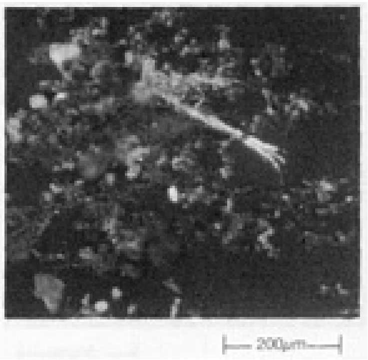


Fig. 8 Particles in oil from reduction gearbox. Red oxides are due to water in oil. 100×

The sample yielded DR ferrograph readings of D_L = 40.6 and D_S = 2.6, giving an unusually high ratio of large to small particles. This size distribution differs drastically from the nearly equal ratio of large to small particles found, for example, in the corrosive wear in a diesel engine. The water compromised the load-carrying capacity of the lubricant, causing a severe wear mode.

As previously discussed, water in oil, at least in concentrations above a few tenths of a percent, can be easily detected by various methods other than ferrography.

Abrasive Wear. Figure 9 is the entry view in which a baseline of wear was established by taking one sample from each of several machines. This ferrogram, from a single machine, shows heavy strings of ferrous wear particles, as well as many large nonmetallic crystalline particles. Compared with a baseline sample, this ferrogram deposit is extremely heavy. Closer examination of Fig. 9 shows that large cutting wear particles dominate the ferrogram. Because there are many nonmetallic crystalline particles, the assumption was made that the cutting wear was caused by abrasive contamination. It was recommended that the oil and oil filter be changed or that a filter bypass rig be used and that the machine be examined for possible means of ingression of contaminants. Another sample from the same machine was submitted a month later, and wear levels had returned to baseline.

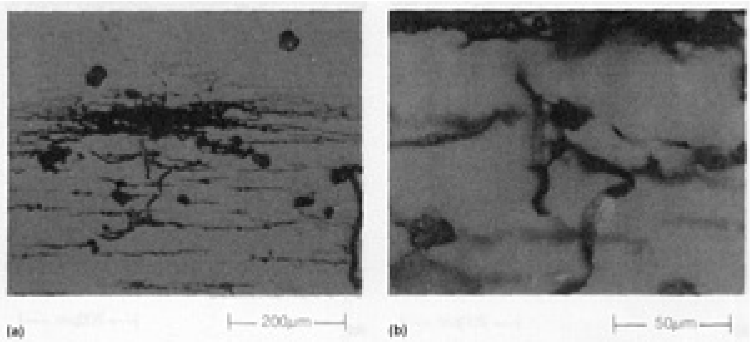


Fig. 9 (a) Entry view of a ferrogram showing a baseline of wear. 100×. (b) Cutting wear caused by abrasion. 400×

A different abrasive wear problem occurred in a refinery moist solvent pump (P-975). The first three samples (10/20/86, 11/15/86, and 11/13/86) from the bearing box of the pump gave DR results that indicated that normal wear conditions existed in the system (Fig. 10). Optical and scanning electron microscopy (SEM) examination of the particles showed dirt particles and a small amount of normal wear debris. At this point, the ferrographic analysis of this machine was no different from that of other unfiltered machines.

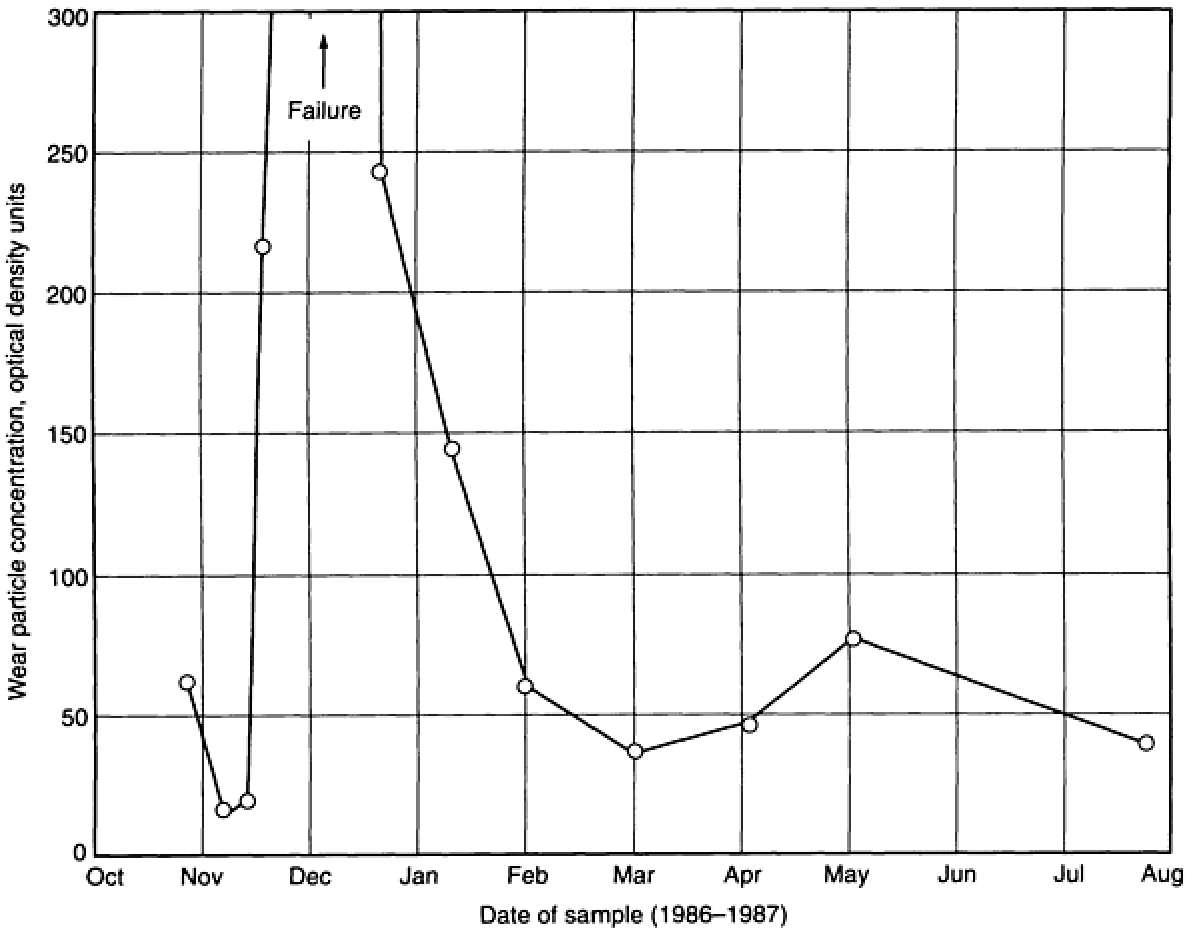


Fig. 10 Wear particle concentration changes in a P-975 moist solvent pump tested over an 11-month period

The fourth sample (11/18/86) gave a 10-fold increase in the large particles (D_L) and a 70-fold increase in severity index, from 105 to 7412. SEM analysis of the particles showed dirt particles and a substantial increase in the wear debris in the form of flakes. Energy-dispersive spectroscopy (EDX) analysis of the flakes indicated that they were steel and babbitt-bearing flakes. X-ray spectrometry of the same samples showed a significant increase in iron, from 35 to 78 ppm (Fig. 11). The combination of ferrography, x-ray spectrometry, and microscopic analysis clearly indicated greatly accelerated rubbing wear caused by dirt contamination. Nevertheless, no action was taken.

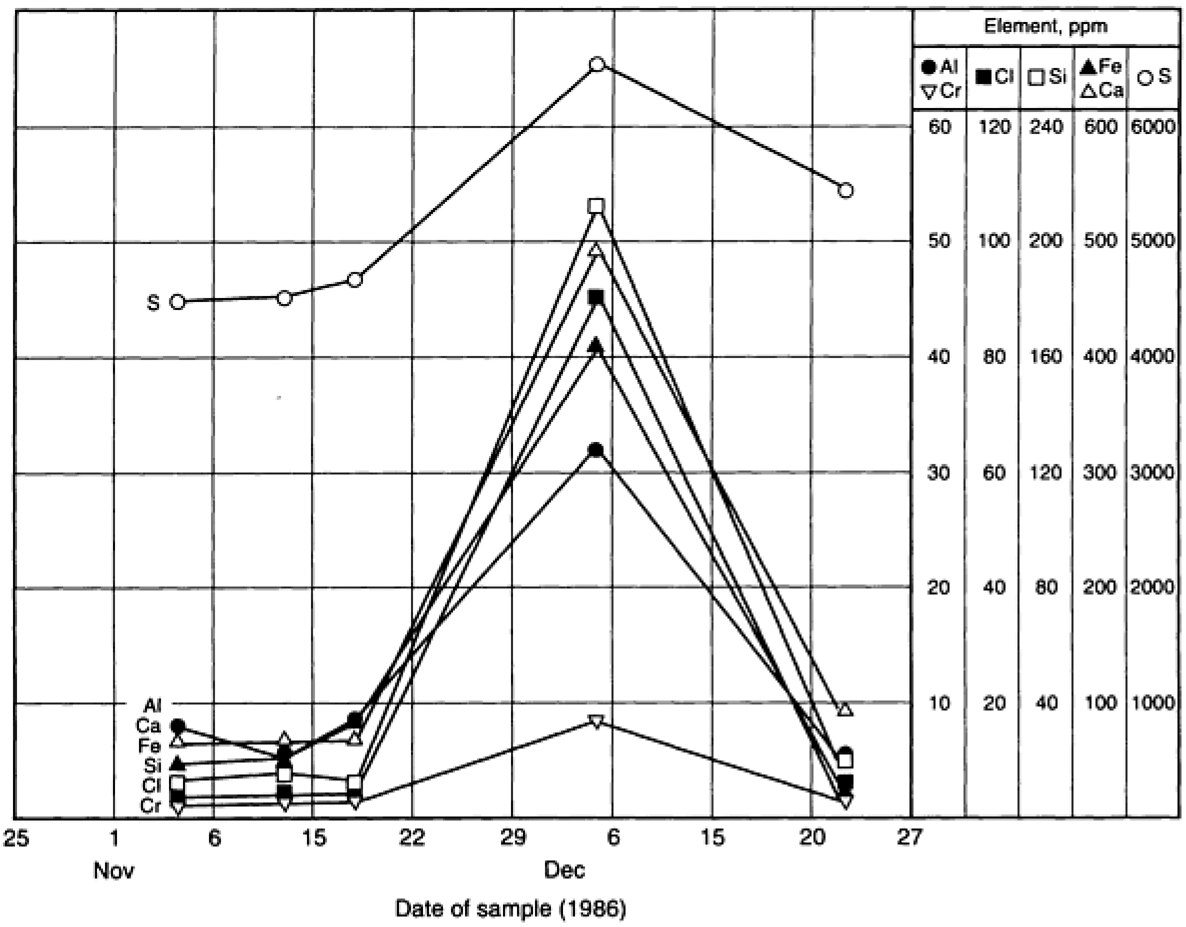


Fig. 11 Spectographic analysis of metals in lubricant samples from a P-975 moist solvent pump before and after failure

On December 5, 1986, a few hours after the fifth sample was drawn, the machine overheated and was shut down before catastrophic failure occurred. The severity index (SI) for this sample was more than 146,000. The 12/5/86 sample was composed of wear debris and lubricant thickened by prolonged heating caused by friction during the abnormal wear mode. The oil was so thick that 10× dilution was necessary before the individual particle could be observed under the microscope. When the equipment was disabled for repairs, maintenance personnel found severe wear on the ball bearings and damage to the race and cage. Subsequently, the damage was repaired, and the pump was cleaned and returned to service.

The 12/22/86 sample showed a substantial decrease in the WPC and SI. SEM analysis showed several strings of wear particles, which were identified as steel and leftover wear debris from the earlier failure. The strings of wear particles were probably break-in wear particles.

The 1/12/87 sample continued to show the break-in wear particles and further reduction in WPC. The February sample showed a return to normal baseline WPC and SI values.

Cast Iron Wear in a Diesel Engine. Figure 12(a) shows the entry deposit (400×) of a ferrogram prepared from a medium-speed marine diesel engine oil sample. The photograph was taken after heat treatment of the ferrogram to 330 °C (625 °F) to distinguish between steel and cast iron. Notice that the particles display the light temper color of cast iron where they do not appear black from oxide formation or from their tortuous shape. The fact that they were heavily oxidized and showed some spots of temper coloring even before heat treatment indicates a high-temperature wear mode, probably caused by inadequate lubrication. Also, the number and size of the particles (many may be classified as severe wear particles) were much greater than normal, implying an abnormal wear mode.

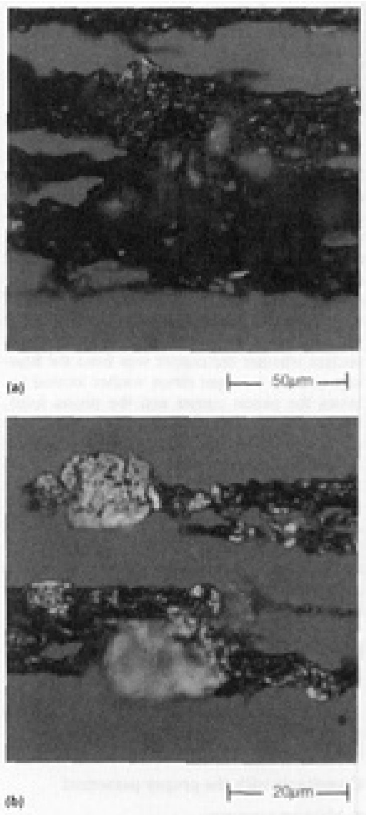


Fig. 12 Entry deposit of a ferrogram from a medium-speed marine diesel engine oil sample. 400 ×. (b) Oxidized cast iron particles with torn shapes. 1000×

Figure 12(b) shows two cast iron severe wear particles at a magnification of 1000×. Even the smaller particles are abnormal in that they are oxidized and have twisted, torn shapes with greater height than in usual for normal rubbing wear platelets. Some weeks after this sample was taken, one of the cylinders froze due to plugging of its oil line.

Corrosive Wear. As previously mentioned, sulfur-containing diesel fuel is corrosive. Alkaline chemical additives are put into diesel engine lubrication oil to neutralize fuel-derived acids as well as to neutralize organic acids, which can form as oil oxidation products. As acid is neutralized, the alkaline additive is consumed. When it is depleted, the engine is exposed to aggressive chemical attack. The result is severe wear, mostly to the piston rings and cylinder liners, although lead in bearings may also be attacked.

Analytically, corrosive wear is readily detected by a low total base number combined with a high wear metals concentration as reported by an emission spectrometer. Total base number (ASTM D 2896) is a measure of the neutralizing capacity of an oil and relates to the amount of alkaline additive in the oil. Corrosive wear can be effectively controlled if the total base number is not allowed to fall below a value of 1.

Corrosive wear is readily detected by ferrography (Fig. 4e), although other oil analysis techniques are also effective. Corrosive wear is indicated by the DR ferrograph by concentration readings that are much higher than the baseline, perhaps 100 times higher. The ratio of large to small particles may nevertheless be very close to 1 because of the absence of many large particles.

Nonferrous Metal Wear in a Marine Engine. An oil sample taken from a 12-cylinder General Motors series 645 engine used for marine propulsion showed a dramatic increase in wear particle concentration compared with a sample taken 2 weeks previously. The DR results were as follows:

Date	Large particles, D_L	Small particles, D_S
July 3, 1980	1.4	1.5
July 7, 1980	39.2	36.2

A ferrogram prepared from the second sample indicated that a piston/cylinder failure was imminent, based on the presence of many severe wear particles as well as many normal rubbing wear particles. These particles were predominantly cast iron, as determined by heat treatment of the ferrogram. Also of concern was the presence of many large, heat-affected copper alloy wear particles. Figure 13 shows one of these particles. At the time of the analysis, it was unclear whether the copper was from the bearings or from a copper thrust washer located between the piston carrier and the piston itself. Eventually, the ship entered a shipyard because of noise in the engine. Two cylinder heads were removed, and signs of excessive wear and scoring of the piston skirt and cylinder liners were observed. Piston thrust washers cannot be visually observed without piston disassembly, but readings can be taken to check clearance to ascertain wear. The copper particles were from a copper thrust that was greatly diminished in size and that showed signs of excessive heat.

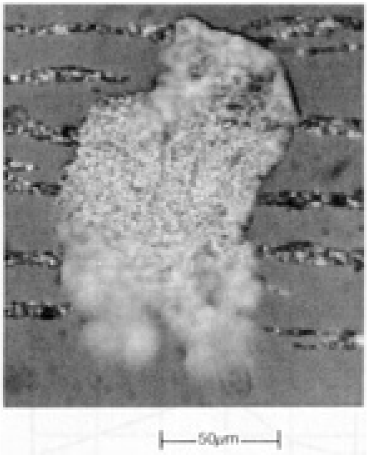


Fig. 13 Copper particles from a piston thrust washer in a marine engine. 400×

Preventive Maintenance Programs

The following recommendations are offered as guidelines for establishing a condition-monitoring program:

Coordinate with the proper personnel

- Machine operators
- Machine maintenance personnel
- Machine service personnel
- Management

Establish personnel involvement

- Document the role of each person
- Use their knowledge to improve the program
- Establish a chain of authority
- Educate all personnel involved

Establish a database before program start-up

- Machine specifications
- Fluid system schematics and specifications
- Operator, service, and repair manuals
- Lubricant type
- Filter specifications
- Wear components, especially material survey data
- Prior service and maintenance reports to identify problem

Establish sampling and sample handling procedures

- Design and document sampling methods, such as in-live valve, modified drain plug, suction tube, and so on
- Ensure cleanliness of sample containers; run some blank ferrograms with filtered oil
- Take sample from system while it is operating, if possible
- Take sample at the same location and machine operating conditions each time
- Coordinate sampling with operator or maintenance personnel
- Provide sampling kits
- Document handling or shipping

Database after start-up

- Program normally should be quantitative;plot DR graphs for each machine
- Initially generate ferrograms for all samples to establish machine signature
- Document all ferrograms with ferrogram analysis sheets
- Store oil sample and ferrograms for possible retrospective analysis

Program administration

- Communications; ensure that results and recommendations reach the appropriate personnel
- Ensure that the role of the analyst within the organization is understood
- Document all substantive work
- Establishment contacts with equipment manufacturer personnel
- Be aware of events in plant and equipment operation that could affect readings temporarily--equipment washing, shutdowns, and so on

Failure Analysis Programs

It is not always understood that post-mortem lubricant screening can significantly speed failure analysis. Difficulty may often arise in obtaining a sample of the lubricant associated with the failure and especially in obtaining a sample of the same lubricant in new condition. The latter is a worthwhile exercise in order to rule out (1) use of the wrong lubricant in the application and (2) unacceptable contamination in the new lubricant. Infrared spectroscopic examination is usually sufficient to rule out these possibilities. Metals analysis often is used to obtain additive metals concentrations and to detect the presence of dirt (Table 3).

The used lubricant should be examined for clues to the cause of machine failure (Ref 41), which commonly include:

- Misalignment or misassembly
- Overheating
- Lack of lubricant, lubricant starvation, inadequate oil changes

- Abrasive or corrosive contamination

Depending on the degree of lubricant or grease decomposition that has occurred, the sample may need to be diluted or filtered. Nevertheless, it is often possible to determine by spectrometric metals and IR analysis the presence of foreign abrasives, corrosive contaminants, water, or fuel. Overheating of the lubricant or old age is readily detected as lubricant oxidation or thermal decomposition via IR analysis. These clues can often be coupled with the results of ferrographic analysis; for example, overheating will result in oxide formation visible as temper colors (black and blue in steel wear debris), whereas predominantly red oxides may confirm rusting caused by water contamination (Ref 47).

Ferrography may also be useful in confirming the type of wear debris. However, it is always preferable to have a preestablished history rather than to draw conclusions from a single sample. In the final failure stages, cutting wear may occur due to roughening and fractured elements (Ref 47). In constructing a story, care must be taken to differentiate changes that caused a problem from those that resulted from the problem.

As an example of a lubricant failure analysis, the used oil from a van was analyzed to determine the cause of severe thickening noted in drawn oil. As shown in the IR spectra (Fig. 14), the used oil contained water but no antifreeze and was very oxidized. Metals analysis, also given in Fig. 14, showed increased additive concentrations, indicating that the oil had been overheated. A high silicon level signified poor air cleaner maintenance and possible engine wear. The overall picture suggested that the engine oil and air cleaner had not been maintained properly, resulting in low oil volumes, oil over-heating, and probably engine wear.

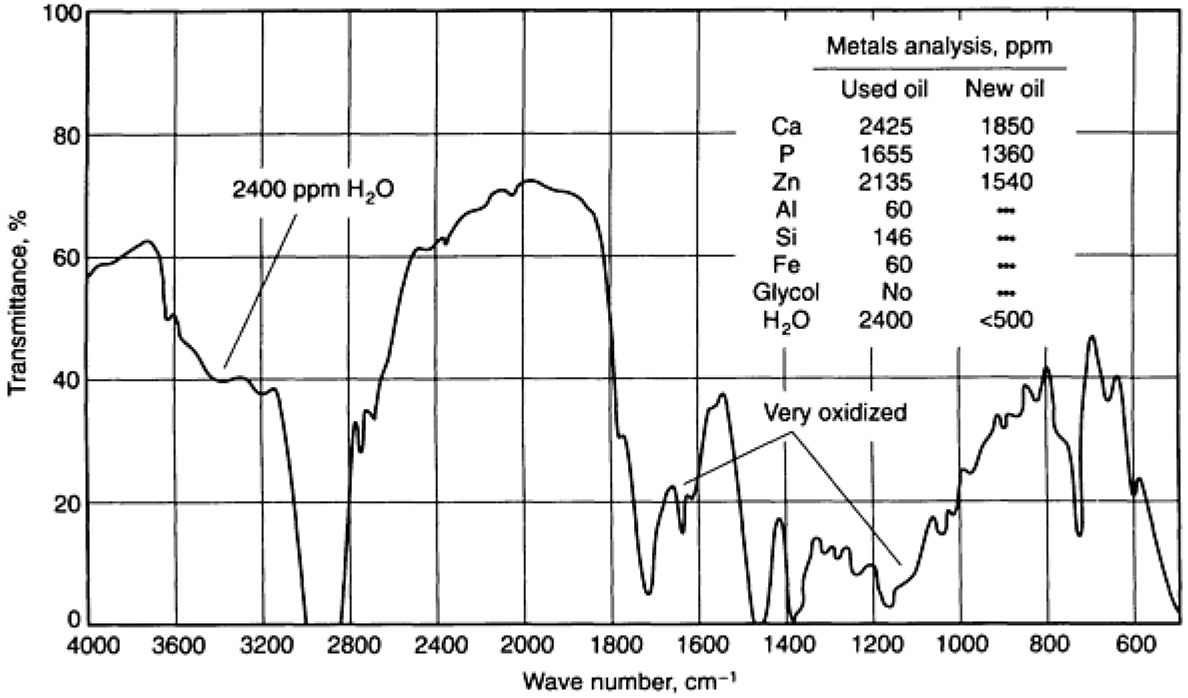


Fig. 14 IR spectroscopy analysis of used engine oil

References

1. A. Decocq, J.J. Brocas, J. DeWindt, and C. Druon, Automatic Wear Metals Control in Diesel Lubricating Oil by ICPES, *Int. Lab.,* June 1987, p 66-72
2. J.S. Stecki and B.T. Kuhnell, Condition Monitoring of Jet Engines, *Lubr. Eng.,* Vol 41 (No. 8), 1985, p 485-493
3. J.S. Stecki, B.T. Kuhnell, D. Equid, and G. McPhee, Ferrographic Analysis of Jet Engine Starter Motors, *Lubr. Eng.,* Vol 43 (No. 1), 1987, p 17-24

4. G.R. Taylor, "Establishment of a Ferrographic Condition-Monitoring Program in a Petrochemical Plant," Preprint No. 82-AM-4C-3, American Society of Lubrication Engineers, 7th Annual Meeting (Cincinnati), 10-13 May 1982
5. F.E. Lockwood, C.M. Comer, J.W. Fu, and F.D. Davidson, Monitoring Wear Particles in Industrial Fluids by Ferrography, *Proceedings Technische, Akademie Esslingen, 6th International Colloquium,* 12-14 Jan 1988
6. J.S. Laskowski, "Sohio's Ferrography Program," p 112-119
7. M. Kurahasi, Applications of Ferrography to Iron- and Steel Making Plant Maintenance, *Maint. Manage, Int.,* Vol 5, 1985, p 269-278
8. A.E. Chichelli, The Role of the Lubrication Engineer in the Steel Industry, Past, Present and Future, *Lubr. Eng.,* Vol 44 (No. 2), 1988, p 147-153
9. S. Corso and R. Adamo, The Application of Ferrography in Monitoring Motor Oils During Engine Development, *Lubr. Eng.,* Vol 45 (No. 9), 1989, p 557-564
10. P.M. Ricca and P.M. Bradshaw, Advanced Features Mark Trans-Alaska Pipeline System's Machinery-Monitoring Program, *Oil Gas. J.,* 20 Aug 1984
11. L.O. Leugner, Use of Sediment Tests and Wear Metals Analyses to Monitor Hydraulic System Condition, *Lubr. Eng.,* Vol 43 (No. 5), 1987, p 365-369
12. L.L. Stavinoha and B.R. Wright, "Spectrometric Analysis of Used Oils," Paper 690776, Society of Automobile Engineers, 1969
13. "Lubricating Oil Analysis for Wear Monitoring," citations from the NTIS database, National Technical Information Service, 1970-Mar 1984
14. "Lubricating Oil Analysis for Wear Monitoring," citations from the NTIS database, National Technical Information Service, Feb 1964-1981
15. R.D. Driver and E.R. Bowen, Wear Particle Equilibrium Measurements and Their Significance, *Proceedings of Joint Conference on Measurements and Standards for Recycled Oil/Systems,* Special Publication 584, National Bureau of Standards, 1980
16. "Standard Test Method for Kinematic Viscosity of Transparent and Opaque Liquids (and the Calculation of Dynamic Viscosity)," D 455, *Annual Book of ASTM Standards,* Vol 5.01, ASTM, 1990, p 170-175
17. "Standard Test Method for Acid and Base Number by Color Indicator Titration," D 974-87, *Annual Book of ASTM Standards,* Vol 5.01, ASTM, 1990, p 330-335
18. "Standard Test Method for Acid Number of Petroleum Products by Potentiometric Titration," D 664-89, *Annual Book of ASTM Standards,* Vol 5.01, ASTM, 1990, p 236-242
19. "Standard Test Method for Water in Liquid Petroleum Products by Karl Fischer Reagent," D 1744-83, *Annual Book of ASTM Standards,* Vol 5.01, ASTM, 1990, p 677-681
20. "Standard Test Method for Detecting Glycol-Base Antifreeze in Used Lubricating Oils," D 2982-85, *Annual Book of ASTM Standards,* ASTM
21. M.P. Granchi, J.A. Biggerstaff, L.J. Hilliard, and P. Grey, Use of a Robot and Flow Injection for Automated Sample Preparation and Analysis of Used Oils by ICP Emission Spectrometry, *Spectrochim. Acta,* Vol 42B (No. 1/2), 1987 p 169-180
22. S.J. Evans and R.J. Klueppel, Analysis of Oil With Inductively Coupled Plasma: Total Process Automation, *Spectrochim. Acta,* Vol 40B (No. 1/2), 1985, p 49-55
23. Y. Liu, A.R. Harding, and D.E. Leyden, Determination of Wear Metals in Oil by X-Ray Spectrometry, *Anal. Chim. Acta,* Vol 180, 1986, p 349-355
24. C.S. Saba, Improving the Wear Metal Detection of Spectrometric Oil Analysis, *Lubr. Eng.,* Vol 46 (No. 5), 1990, p 310-316
25. R.E. Kaufmann, Technique to Improve the Wear Metal Detection Capabilities of Rotating Disk Electrode-Emission Spectrometers, *Lubr. Eng.,* Vol 46 (No. 3), 1990, p 173-180
26. Y. Nakamura, T. Etoh, Y. Noto, and Y. Murai, "Determination of Wear Metals in Lubricating Oils by Inductively Coupled Plasma Emission Spectrometry," Translation No. 11522, Health and Safety Executive

Translation Service.

27. R.E. Kaufmann, Particle Size and Composition Analyses of Wear Debris Using Atomic Emission Spectrometry, *Lubr. Eng.,* Vol 45 (No. 3), 1989, p 147-153
28. A. Bond and D. Kittinger, "Spectrometric Oil Analysis for the Detection of Incipient Turbo Engine Failures," AD472771, U.S. Department of Commerce, 1965
29. K.N. Subramanian, Spectrographic Examination of Sump Oil as Predictive Tool in the Maintenance of Diesel Engines, *Corros. Maint.,* Apr-June 1981, p 127-129
30. Perkin-Elmer INFRARED Bulletin 20, IRB-20, Perkin-Elmer, Norwalk, CT
31. J. Poley, Oil Analysis for Monitoring Hydraulic Oil Systems, A Step-Stage Approach (c), *Lubr. Eng.,* Vol 46 (No. 1), 1990, p 41-47
32. L.O. Leugner, The Use of Oil Contamination Testing Combined With Improved Filtration Will Optimize Both Lubricant and Equipment Life, *Lubr. Eng.,* Vol 45 (No. 10), 1989, p 618-624
33. R. Cooper, Wear Debris Monitoring of Rolling Bearings, *Br. J. Nondestr. Test.,* Mar 1983, p 75-83
34. P.S. Baur, Ferrography: Machinery-Wear Analysis With a Predictable Future, *Power Mag.,* 1982
35. U.S. Patent No. 4047814
36. E.R. Bowen and W. Seifert, "Ferrography--A New Tool for Analyzing Wear Conditions," paper presented at Fluid Power Testing Symposium (Milwaukee), 16-18 Aug 1976
37. D.P. Anderson and R.D. Driver, Equilibrium Particle Concentration in Engine Oil, *Wear,* Vol 56, 1979, p 415-419
38. W.W. Seifert and V.C. Westcott, A Method for the Study of Wear Particles in Lubricating Oil, *Wear,* Vol 21, 1972, p 27-42
39. A.A. Reda, R. Bowen, and V.C. Westcott, Characteristics of Particles Generated at the Interface Between Sliding Steel Surfaces, *Wear,* Vol 34, 1975, p 261-273
40. R.H. Rotondi, Ferrography--The Technology of Wear Particle Analysis, *Heavy Duty Equip. Manage. Maint.,* May 1980
41. D. Scott, Debris Examination--A Prognostic Approach to Failure Prevention, *Wear,* Vol 34, 1975, p 15-22
42. *Wear Particle Atlas,* Predict Technologies, p 31
43. D.W. Dudley, *Handbook of Practical Gear Design,* McGraw-Hill, 1984, p 2.20
44. M.V. Hofman and J.H. Johnson, The Development of Ferrography as a Laboratory Wear Measurement Method for the Study of Engine Operating Conditions on Diesel Engine Wear, *Wear,* Vol 44, 1977, p 183-199
45. S. Corso and R. Adamo, "Incipient Scuffing Detection by Ferrography in a Diesel Valve Train System, SAE Paper 852124, Society of Automotive Engineers
46. "Wear in Fluid Power Systems," Final Report to the Office of Naval Research, Contract No. N00014-75-C-1157, Fluid Power Research Center, June 1979
47. E.R. Bowen and R.D. Driver, Water Particle Equilibrium Measurements and Their Significance, *Proceedings Joint Conference on Measurements and Standards,* Special Publication 584, National Bureau of Standards, 1979

Motor-Current Signature Analysis

D.M. Eissenberg and H.D. Haynes, Oak Ridge National Laboratory

Introduction

FRICTION-GENERATED WEAR is a major cause of degradation of mechanical equipment. Wear can result from the normal operation of the equipment, or it can be accelerated because of operation under unanticipated severe conditions, or it can result from other degradations of the equipment that exacerbate the wear. The diagnosis and correction of wear, particularly accelerated wear, before it causes equipment failure can lead to a significant reduction in overall maintenance costs. Depending on the application, the detection and trend tracking of wear can also improve safety and reliability, as well as reduce unscheduled outages.

The key to successful early diagnosis and correction of wear in various types of equipment is the availability of cost-effective nonintrusive techniques for sensing, monitoring, and diagnosing either the causes or the extent of the wear. This article describes a new diagnostic monitoring technique, motor-current signature analysis (MCSA), which provides early diagnostic capabilities for detecting wear for the subset of mechanical equipment that is driven by electric motors.

The MCSA technique was developed by Oak Ridge National Laboratory (ORNL) to provide an improved means of determining the condition of motor-operated valves (MOVs) widely used as isolation or control valves in nuclear power plant safety systems (Ref 1). MCSA has been successfully tested on MOVs in nuclear power plants and has been demonstrated at the Electric Power Research Institute (EPRI) Monitoring and Diagnostic Center. Experience with the application of MCSA to motor-operated valves has shown that the technique detects MOV wear as well as other types of degradation.

Described below are the principles and equipment associated with MCSA, followed by specific examples of its application to the measurement of friction and wear of motor-operated valves.

Acknowledgement

The research described in this article was sponsored by the Nuclear Regulatory Commission Nuclear Plant Research Program and by the Oak Ridge National Laboratory Advanced Diagnostic Engineering Research and Development Center. Oak Ridge National Laboratory is operated by Martin Marietta Energy Systems, Inc. under contract DE-AC05-84OR21400 with the U.S. Department of Energy.

Operating Principles

The technical basis of MCSA is the observation that an electric motor driving a mechanical load acts not only as a transducer, in the sense of converting electric power to mechanical power, but also in the reverse sense, as a permanently available and easily accessible transducer that converts the mechanical load features of the driven device into an electrical signature. It is noted that for most types of electric motors, the relationship between current and output load is not linear, and is dependent on the supply voltage. However, that relationship is a characteristic of the motor, and characteristic curves can be obtained from the manufacturer for each motor over a range of input voltages.

The electric-current signature can be decomposed into two elements:

- The mean value of the current supplied to the motor and its (gradual) change with time. In the case of ac-powered motors, the mean value is the average root-mean-square (RMS) value
- The instantaneous fluctuations in the motor current. The amplitudes of the fluctuations are in general small, when compared to the mean value of the current

Both types of motor-current information can be useful, although in different ways, in diagnosing friction and wear. The mean value of the electric current provides a direct measure of the mean value of the total mechanical load of the driven device, and thus is a measure of the friction load within the device. In addition, gradual changes in the mean value of the motor current may reflect changes in the coefficient of friction that are due, for example, to the roughening of surfaces in sliding contact within the driven device.

Instantaneous motor-current fluctuations appear superimposed on the mean current. They can be either periodic or nonperiodic. In each case, the characteristics of the motor-current noise reflect the fluctuations in the mechanical load that the motor is driving. Nonperiodic load fluctuations (transients) occur, for example, during start-up of the motor-driven device, as well as when the device is subjected to rapid changes in load. In the case of MOVs, this occurs during valve seating or unseating.

Periodic fluctuations occur when there are rotating elements in the driven equipment. Gears, bearings, pistons, multivane impellers, and rotating seals are all sources of periodic load fluctuations.

Motor-current signatures are analyzed in both the time domain and the frequency domain using commercially available equipment, including PC-based programs. The time domain provides is optimum approach for detecting and monitoring both the mean value and the transients. The time of occurrence and the shape of the transient both provide useful information. The frequency-domain signatures are used to analyze periodic fluctuations. The frequency of rotation of each rotating element of the load usually can be readily identified, and the harmonics associated with that frequency are used to determine the extent of service wear of that rotating element.

MCSA Equipment

The application of motor-current signature analysis involves three types of equipment: signal acquisition and recording, signal processing, and signal analysis. Signal acquisition is accomplished using conventional nonintrusive electric-current sensors. These include a current transformer (split jaw or permanently installed coil) for ac-powered motor-driven equipment and a Hall-effect probe for dc-powered motor-driven equipment. It is usually necessary to place the current-measuring equipment on only one of the power leads. The current sensor can be placed anywhere along the power leads, from the motor itself to the motor control center or other circuit branch point or power transformer. The resulting current signal is amplified prior to further processing. The amplifier must be one that does not introduce significant signal distortion over its expected dynamic range.

For ac-powered motor-driven equipment, it is generally necessary to demodulate and filter the signal. This is because the 60-Hz frequency component and its harmonics will tend to make further analysis more difficult, particularly when signals close to 60 Hz or its harmonics are present.

The demodulation can be performed using several techniques. One that has been found effective is RMS-to-dc conversion. Generally, a low-pass filter is used after the RMS-to-dc conversion, although in some special cases a notch filter and band-pass filter have been found useful. A functional block diagram of a motor-current signal conditioning circuit is shown in Fig. 1.

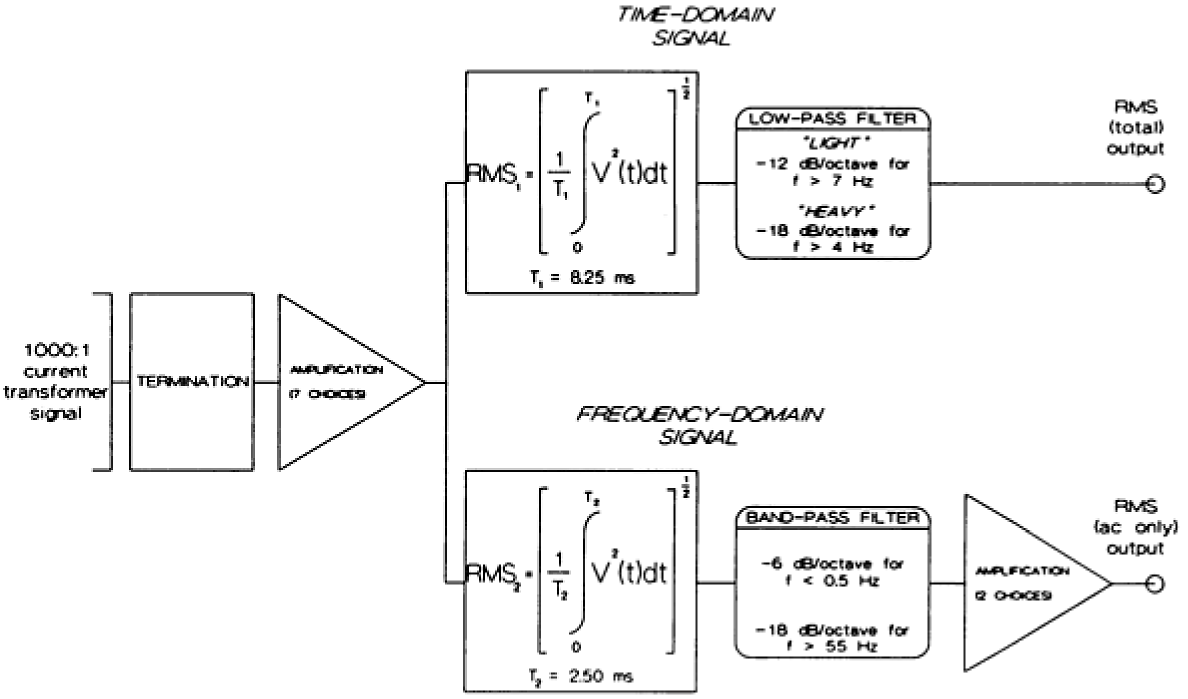


Fig. 1 Functional block diagram of a motor-current signal conditioning circuit used for ac-driven motor applications

Two separate output signals are developed from the amplified, demodulated, and filtered signal: one is optimized for time-domain analysis and the other is optimized for frequency-domain analysis. The signal conditioned for time-domain analysis is digitized, and the signature is displayed on a digital oscilloscope. The signal that is optimized for frequency-domain analysis is digitized and fed to a fast Fourier transform (FFT) analyzer for display of the frequency-domain signature (Fig. 2). Instead of using separate instruments, it is convenient to use a commercially available computer program that combines both oscilloscope and FFT display functions using a personal computer.

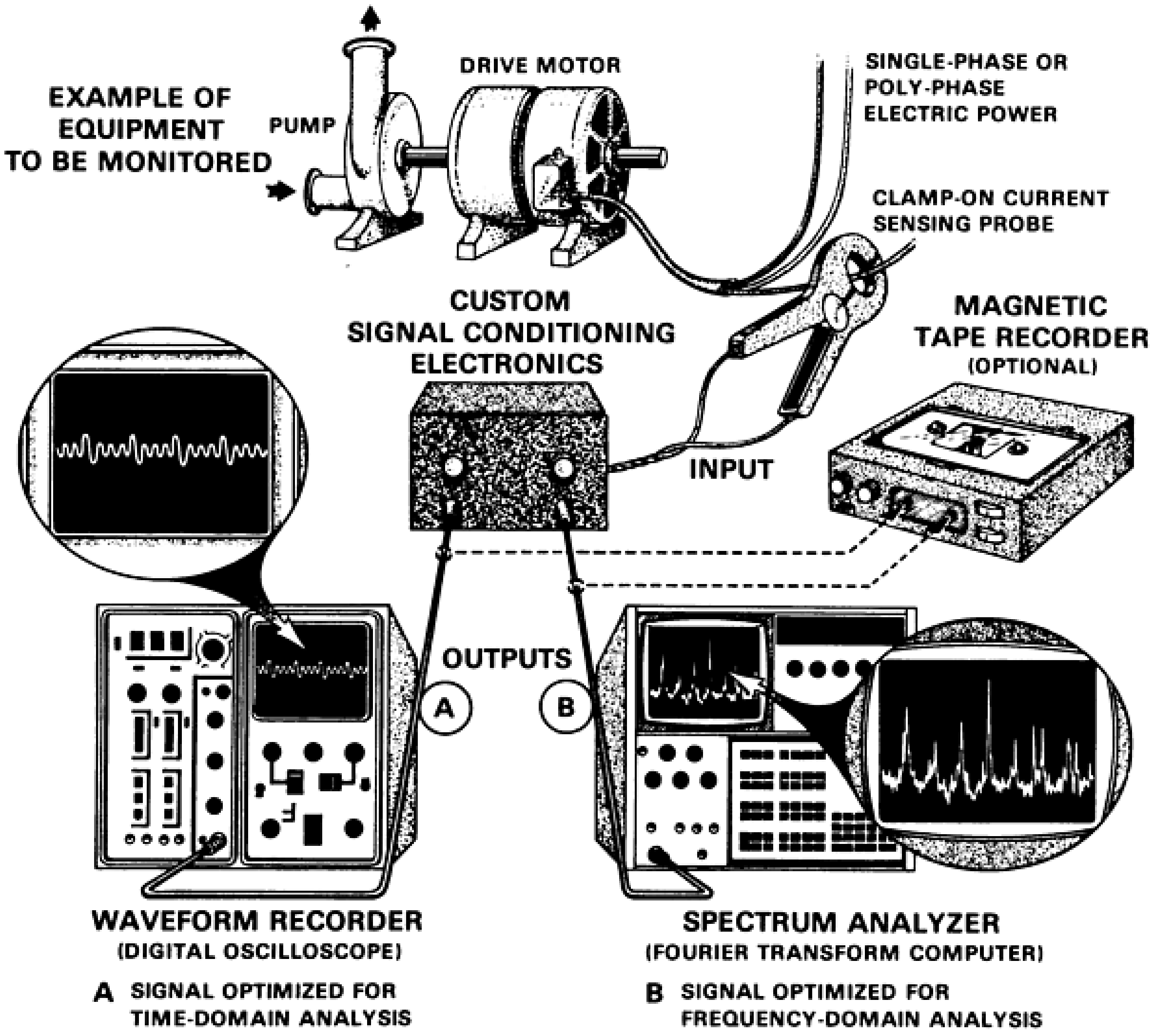


Fig. 2 Motor-current signature analysis (MCSA) method

Application to Motor-Operated Valves

ORNL has carried out extensive laboratory and field investigations to evaluate the capability of MCSA for determining the operational readiness of MOVs used in the safety systems of nuclear power plants. Operational readiness is the ability to function under all anticipated operating conditions, including those associated with accident mitigation and recovery. Readiness is affected by many factors, including improper installation, improper setup, and the presence of defects (degradations) that result from operations under the extreme conditions that occur during or following an accident. Investigations show that MCSA can provide useful wear information, as illustrated by the examples below.

Stem-Nut Wear. During a disassembly inspection, the stem nut from a motor-operated 150 mm (6 in.) globe valve was discovered to have suffered extensive thread damage (Fig. 3). Investigations revealed that the root cause of this damage was a lower grease seal spring that had become loose during an earlier, improperly completed installation of the drive sleeve and had entered the stem-nut thread region. During subsequent valve actuations, the spring had become wedged between the stem nut and the stem, causing the accelerated wear of the stem nut.

Fig. 3 Valve stem nut showing wear of threads

The damage affected the motor-current signatures. In the time-domain signature, there was an increase in the time interval between the motor start-up and the initial stem movement (pickup time), which reflected the increase in clearance between the stem nut and the stem threads.

A precise measurement of the pickup time was made for signatures obtained from the 45 valve cycles during which stem-nut damage occurred and also after a new stem nut was installed. A plot of the pickup time for open-to-close valve strokes obtained during the 45-cycle sequence is shown in Fig. 4. The relationship between increases in this time and increases in stem clearance was calculated based on the known stem speed, 1.7 mm/s (67 mils/s). Thus, the approximately 0.2 s increase in time differential corresponds to an increased clearance of 330 m (13 mils) that was due to the accelerated wear.

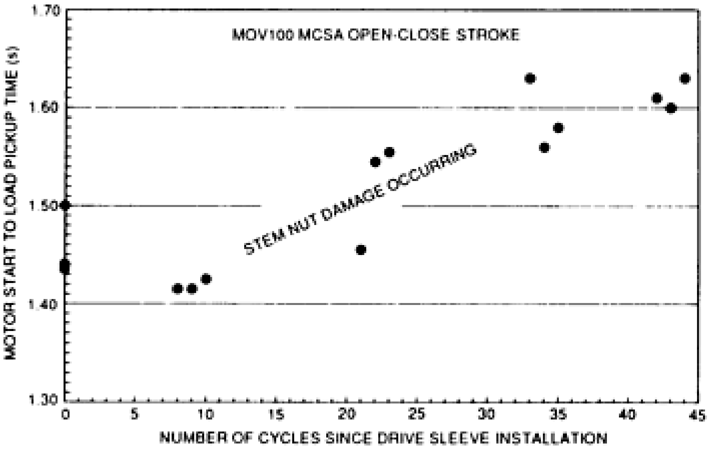


Fig. 4 Influence of valve stem nut wear on load pickup time for 150 mm (6 in.) motor-operated globe valve

A review of historical data for that valve revealed a longer-term trend in the pickup time that had preceded the incident cited above. This is shown in Fig. 5, which includes data collected over the last 650 valve cycles. It is seen that even without the accelerated wear resulting from the broken spring, there had been a gradual trend of increasing clearance. Also shown is the pickup time that resulted after installation of the new stem nut. The results are consistent with the conclusion that the pickup time as measured by the motor-current signature is a precise measure of stem-nut wear.

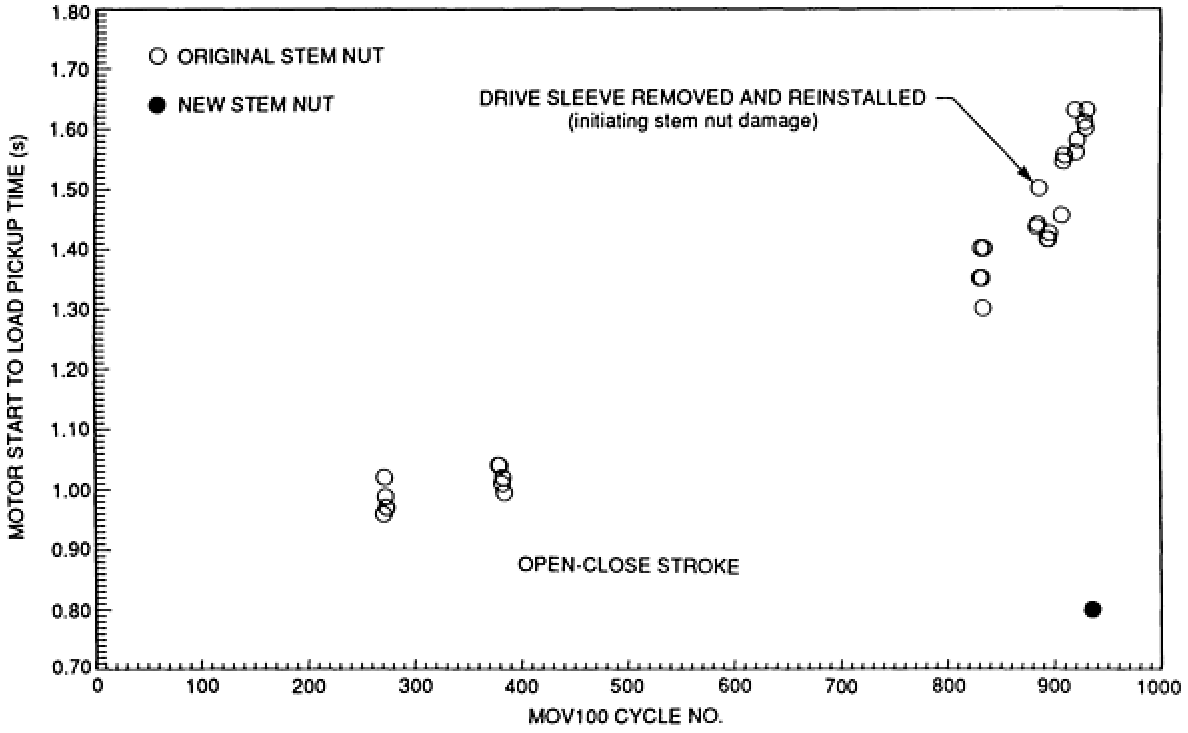


Fig. 5 Historical measurements of load pickup time for 150 mm (6 in.) motor-operated globe valve

An examination of the frequency-domain signatures of the valve motor current during mid-stroke provided additional evidence that accelerated stem-nut wear was occurring. Figure 6 presents two 10- to 20-Hz spectra of motor current. One was acquired before the improper drive sleeve installation and one soon after the drive sleeve was installed. Sidebands, located around the worm-gear tooth-meshing frequency at spacings equal to the stem-nut rotation frequency, approximately doubled in average amplitude soon after the drive sleeve was installed, during the time when stem-nut damage most likely occurred.

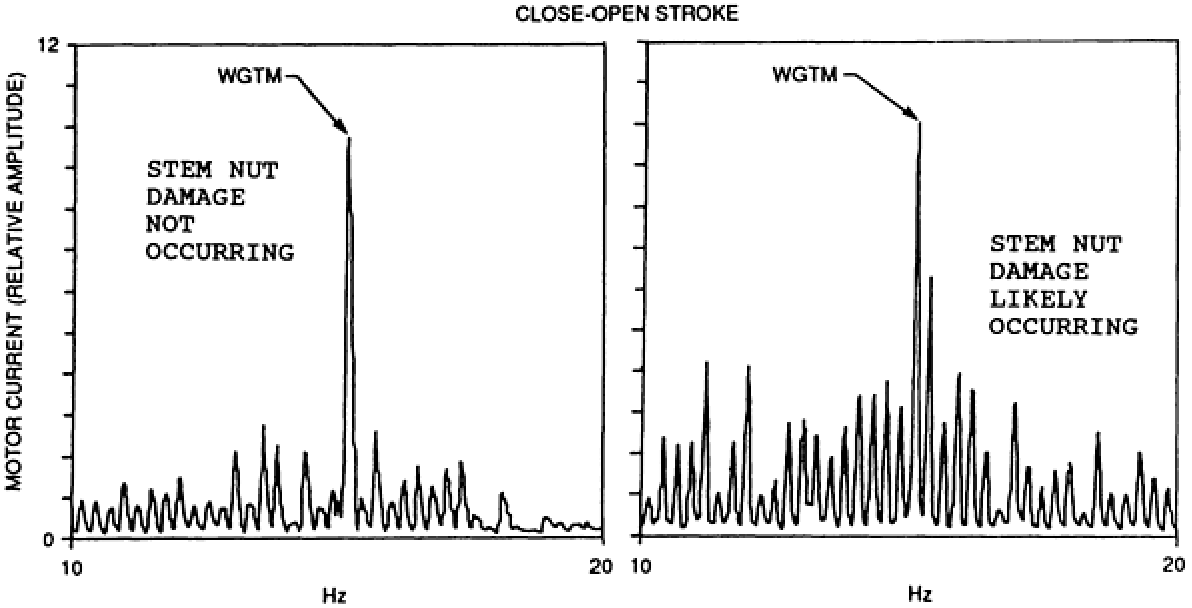


Fig. 6 Influence of valve stem-nut damage on frequency spectrum of demodulated motor-current signals from

150 mm (6 in.) motor-operated globe valve

These sidebands were observed to return to normal amplitude levels by the end of the 45 cycles. This suggested that the sidebands reflected a dynamic occurrence (stem-nut galling in progress), rather than a residual condition (after stem-nut damage had occurred). The diagnostic information acquired from this method thus complemented the differential time measurement described previously as an indicator of residual stem-nut wear.

Degraded Worm and Worm-Gear Lubrication. The worm and worm gear of a motor operator are normally lubricated with grease in order to reduce friction and wear. In one instance, a valve was actuated inadvertently without sufficient lubrication in this area. This occurred because the valve was used for training and was frequently being assembled and disassembled. Motor-current signatures were obtained during valve actuation with no lubrication under two conditions of load: the as-found condition when the valve packing was loose, and the condition after the packing was tightened. The results (Fig. 7) indicate a normal signature for the loose packing, but a high motor-load condition when the packing was tightened moderately. Because the high loads occurred in the drive train between the motor and the torque switch, the normal torque switch cutout of the motor did not occur and the motor stalled. When the worm and worm gear were lubricated, actuations with both loose and tight packing yielded normal motor-current signatures, as shown in Fig. 7.

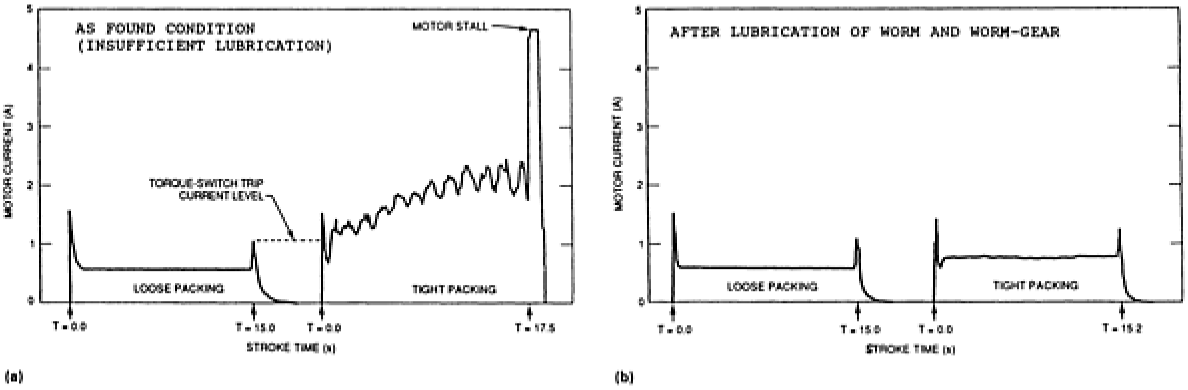


Fig. 7 Open-to-close stroke actuations of a 75 mm (3 in.) motor-operated valve before and after lubricating the worm and worm gear. (a) As-found condition (insufficient lubrication). (b) After lubrication of worm and worm gear

Gear-Tooth Wear. The detection of gear-tooth wear is illustrated by the time-domain plots of Fig. 8. The uppermost plot shows a 10 s midstroke portion of the motor-current signature at the start of the test, that is, the baseline condition with no gear-tooth wear. The ordinate scale has been expanded and offset from zero to show clearly the cyclic variations in motor current that result from motor slip and the engagement of individual teeth of the MOV worm gear with the worm (which occurs with each revolution of the worm, that is, each 65.16 ms for this particular motor operation).

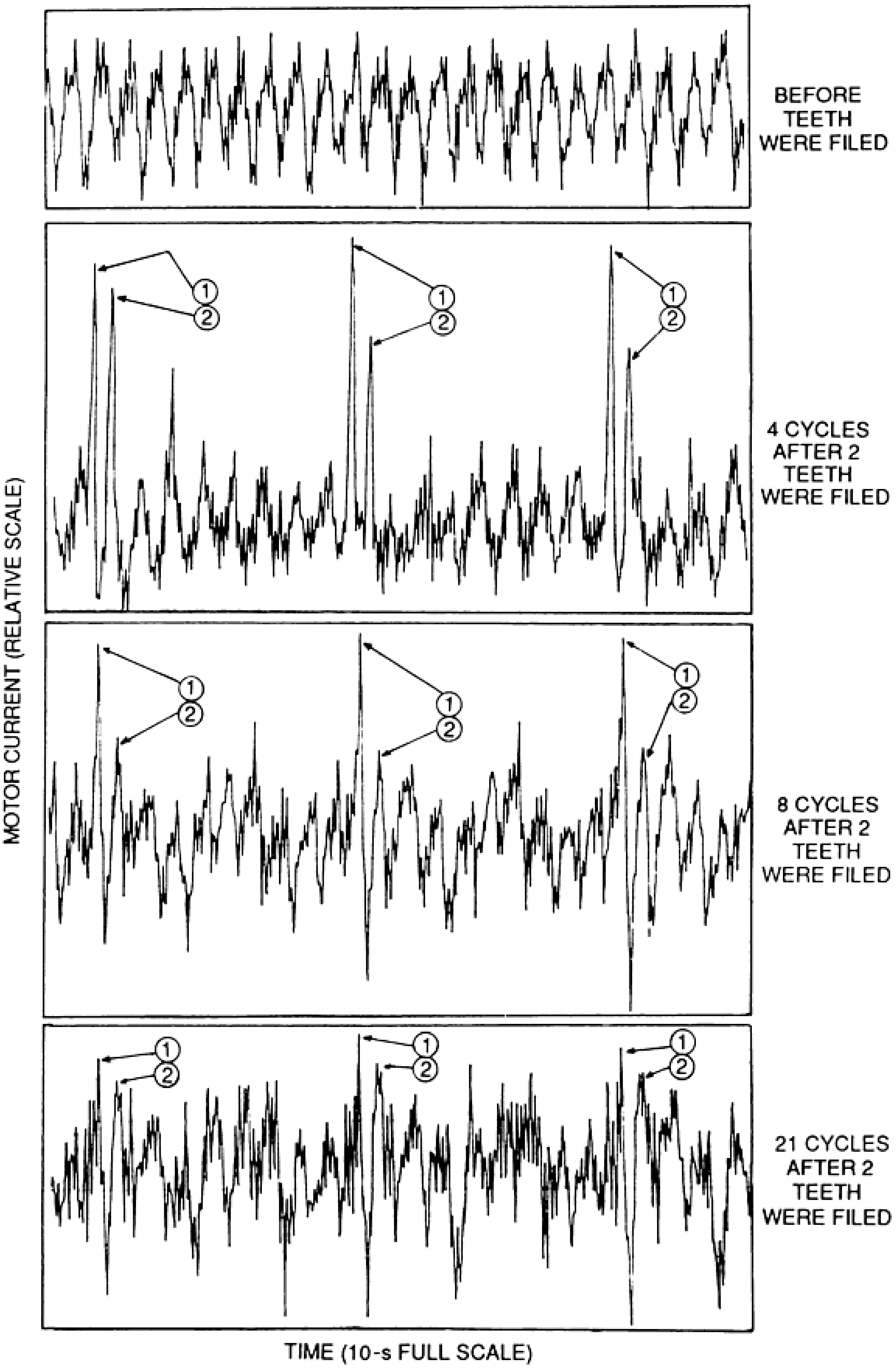


Fig. 8 Impact of worm-gear tooth wear on motor-current signatures; frequency, 2,33 Hz; time period, 0.429 s

For the purpose of demonstrating the capability of the motor-current signature to detect this type of wear, the worm gear was removed from the operator and the contact surfaces of two non-adjacent teeth were filed to both remove metal and roughen the surfaces. Approximately 165 m (6.5 mils) were removed from one tooth and 125 m (5.0 mils) from the other, later-engaging tooth. (In each case, the material was removed symmetrically from both sides of the tooth and from across the entire tooth width, in order to maintain tooth surface angles.)

After reinstallation of the defective gear, the operator was actuated repeatedly, giving the results shown in the lower three traces of Fig. 8. The presence of two strong peaks corresponds to the successive meshing of the two altered gear teeth with the worm. Note that the more heavily filed tooth gave the larger transient and that the pair of peaks recur each 3.7 s, which is the time required for one complete revolution of the worm gear. The peaks became less pronounced as the operator was stroked repeatedly, presumably the result of gradual "wearing in" of the implanted defects.

Stem Taper Effect on Packing Friction. When the motor-current signature of a 150 mm (6 in.) globe valve was first examined, the current level decreased during the open-to-close valve stroke and increased in a symmetrical fashion during the close-to-open stroke, as shown by Fig. 9. Because of the symmetrical and smooth nature of this reproducible motor-current signature feature, the current level behavior was assumed to be an indication of a valve stem anomaly. A close inspection of the valve stem after disassembly provided the reason for this signature characteristic. It was found that the diameter of the stem varied within the region that was in contact with the stem packing during valve actuations, as indicated by Fig. 10.

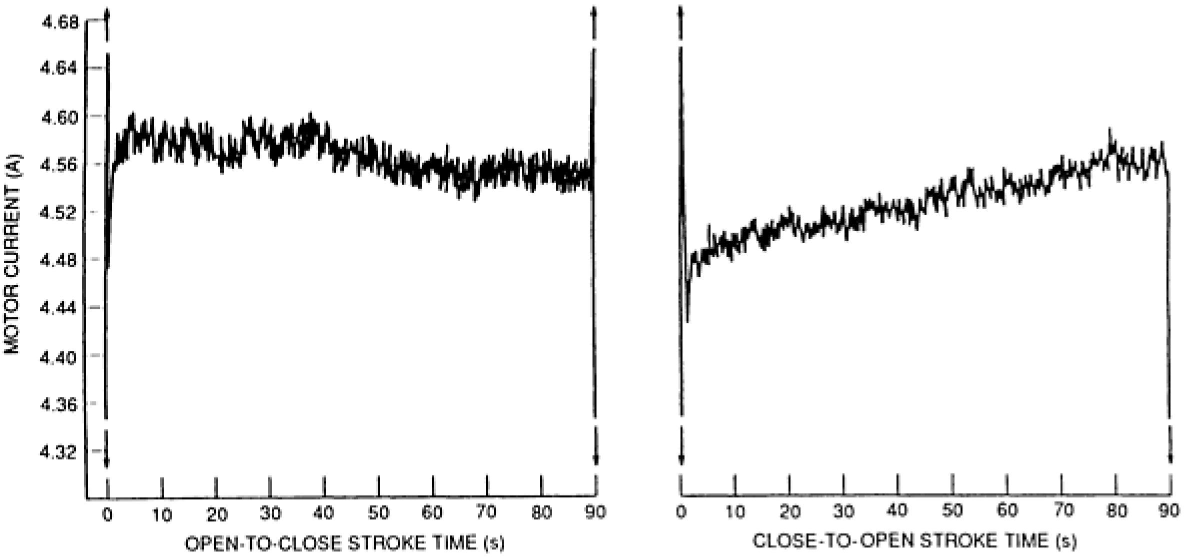


Fig. 9 Motor-current signatures for 150 mm (6 in.) globe valve showing influence of valve stem taper

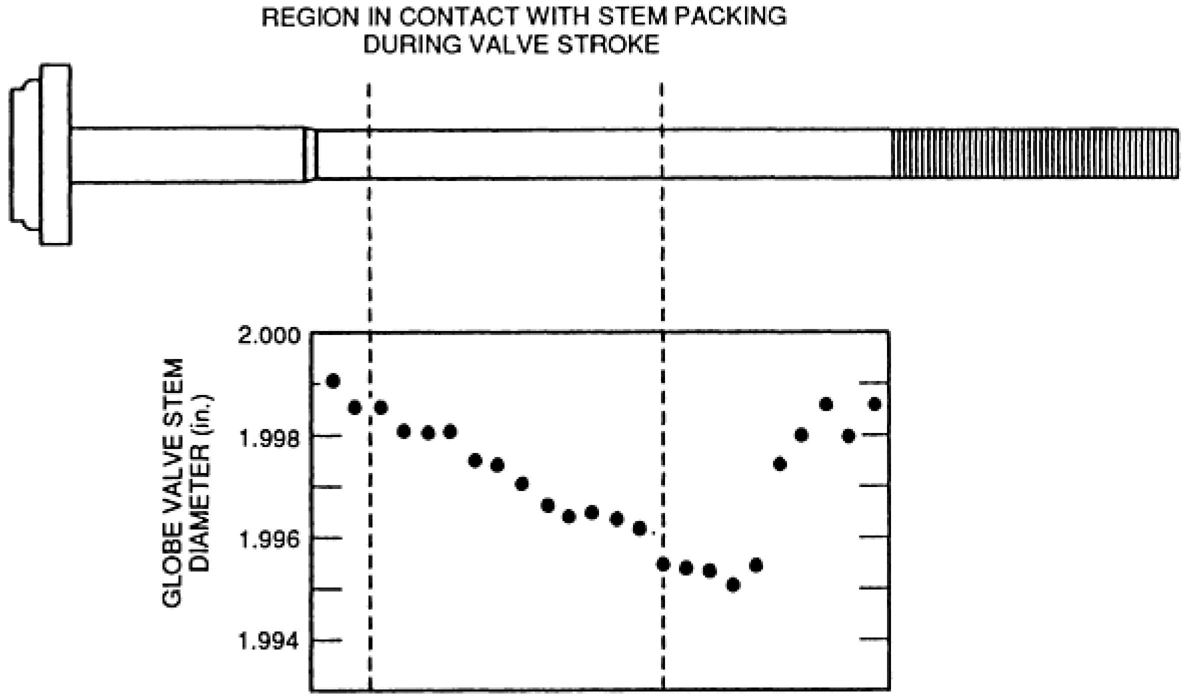


Fig. 10 Tapered valve stem of 150 mm (6 in.) globe valve illustrated in Fig. 9

As the valve was actuated in the open-to-close direction, the average stem diameter within the packing gland decreased. This reduced the packing friction load and therefore reduced the load driven by the motor. Conversely, during the close-to-open stroke, the average stem diameter within the packing gland increased, and the loads driven by the motor increased during the valve stroke. The current rise during the close-to-open stroke was greater than the decrease in the aforementioned open-to-close stroke. This may have reflected a relaxation of the packing load in the interval following completion of the close-to-open stroke.

Gate-Valve Unseating. Wedge-gate valves are seated by the wedging action of the tapered disk as it is forced between two tapered seat rings. Both the seating and the unseating transients can be seen in motor-current signatures. A typical normal unseating transient appears in Fig. 11(a).

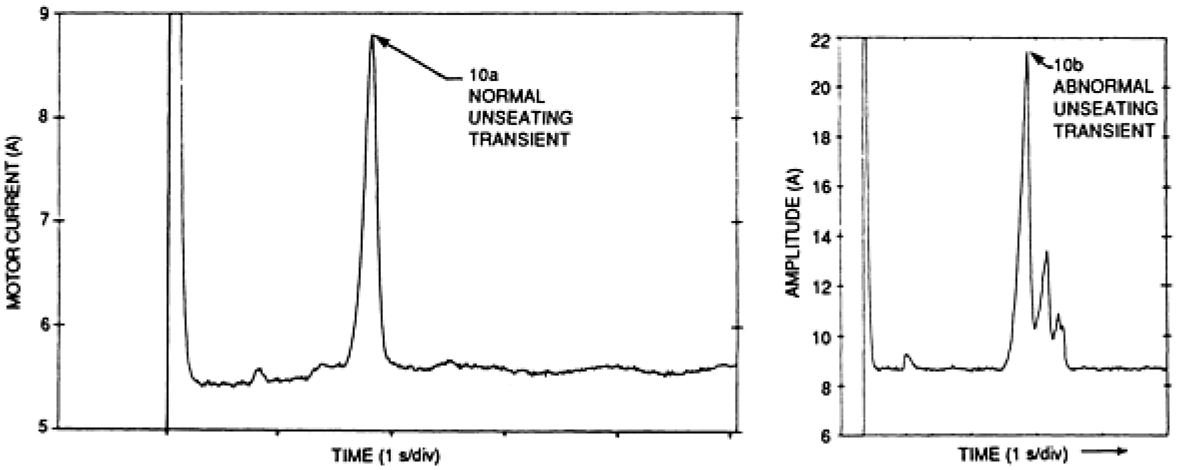


Fig. 11 Normal and abnormal unseating transients observed on two different motor-operated gate valves

In tests carried out at a nuclear power plant, the unseating transient of one of the valves appeared as shown in Fig. 11(b). It appears that a stick-slip condition occurred in the unwedging of that valve. Because of operating conditions, it was not possible to disassemble the valve to determine the cause of this behavior.

Actuation of Wedge-Gate Valves under Differential Pressure. When a wedge-gate valve is opened with a pressure drop across the gate or when the valve is closed while there is a flow of water through the valve such that a pressure drop builds up as the valve is closed, there is a frictional load between the gate and stationary seat or guide surface that is proportional to the pressure drop across the sliding gate. For some valves, this can result in excessive wear, including galling of the affected surfaces and an inability of the motor operator to open or to fully close the valve.

Motor-current signatures were obtained during actuations of a motor-operated gate valve with and without differential pressure. Shown in Fig. 12(a) and 12(b) are time-domain motor-current signatures obtained while the motor-operated gate valve was closed under two different conditions. In Fig. 12(a), the valve was closed while nonflowing internal fluid was at a temperature of 275 °C (530 °F) and a pressure of 7 MPa (1 ksi). The pressure drop across the gate during closure was zero. In Fig. 12(b), the motor-current signature is shown when the same valve was closed at the same fluid temperature, but with flow through the valve such that, at closure, there was a pressure drop across the gate of 7 MPa (1 ksi).

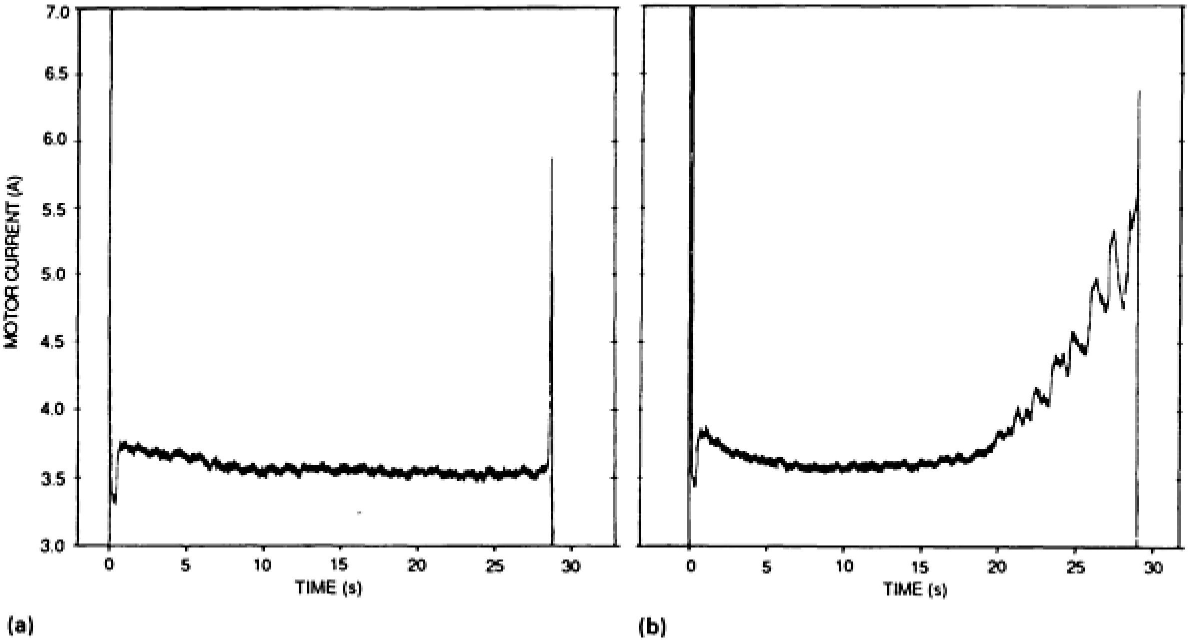


Fig. 12 Open-to-close stroke motor-current signatures for a gate valve. (a) Before blowdown. (b) During high-flow (blowdown) test

The following observations were noted. In Fig. 12(a), there was a virtually constant motor current during the stroke. In Fig. 12(b), the mean value of motor current increased sharply during the latter portion of the stroke as the gate blocked the flow, as would be expected as a result of the increased friction load between the gate and seat surfaces. In addition, there was a periodic variation in the motor current (also picked up by a direct-reading strain gage) that was due to rubbing of the lower edge of the gate against the lower edge of the seat. In some wedge-gate valves, the clearance between the gate and guide surfaces is such that the gate can tilt and rub the seat as noted above. Subsequent disassembly and inspection of the rubbing surfaces confirmed that excessive wear had occurred.

Future Trends

The above-cited and other data support the conclusion that MCSA is a useful tool for monitoring the mechanical and electrical condition of MOVs, particularly in relation of their operational readiness. Experience with motor-driven machinery other than MOVs, though limited, strongly suggests that MCSA is equally applicable to monitoring present conditions and to diagnosing impending trouble in a wide variety of consumer and industrial equipment. MCSA has a number of inherent strengths, the most notable being that it:

- Provides nonintrusive monitoring capability at a location remote from the equipment
- Provides degradation and diagnostic information comparable to conventional instrumentation (accelerometers, for example) but without the attendant disadvantages of added sensors and signal cables
- Offers high sensitivity to a variety of mechanical disorders affecting operational readiness
- Offers means for separating one form of disorder from another (selectivity)
- Can be performed rapidly and as frequently as desired by relatively unskilled personnel using portable, inexpensive equipment
- Is equally applicable to high-powered and fractional horsepower machines, driven by either ac or dc motors

Reference

1. H.D. Haynes *et al.,* Motor Current Signature Analysis Method for Diagnosing Motor Operated Devices, U.S. Patent 4,965,513, 23 Oct 1990

Selected References

- H.D. Haynes, Aging and Service Wear of Electric Motor-Operated Valves Used in Engineered Safety-Feature Systems of Nuclear Power Plants, *Vol. II: Aging Assessments and Monitoring Method Evaluations,* NUREG/CR-4234, Oak Ridge National Laboratory

Radionuclide Methods

Charles C. Blatchley, Spire Corporation

Introduction

GAMMA-EMITTING RADIONUCLIDES have been widely used for over 40 years to measure wear in engines, high-temperature environments, and lubrication testing. Their unique ability to provide on-line *in-situ* readings, both effectively and in real time, has made approaches based on them valuable for many engineering problems in either system design or lubricant evaluation. Extreme precision also makes these methods ideal for basic tribological research, particularly for relating surface loss, which is a critical effect of mechanical interactions, to other tribological aspects, such as friction.

Radionuclides Applied to Tribology

As material loss occurs, gamma-emitting material is removed, along with the original surface material. A gamma detector measures either the decrease in intensity from the activated part (marker approach) or the increase in debris collected in a filter or trap (tracer approach).

Tracers. First applications of radionuclides in wear testing used reactor-generated wear debris tracers as a diagnostic tool in 1949 (Ref 1). Tracer use became common in diesel engine testing through the 1960s, particularly in locomotives (Ref 2, 3, 4, 5, 6, 7, 8, 9). They are still often applied to engines, either where shielding prevents the effective use of other approaches or where timing of wear onset is more important than quantity removed.

Markers. Surface layer activation (SLA) by accelerator bombardment, rather than neutron irradiation, allows the creation of shallow surface markers that can be monitored remotely for decreases in activity. Although any activated material can be used for tracing (debris activity increases), direct monitoring of a marker allows much greater precision,

particularly if the marker is sufficiently shallow. Marker monitoring requires much less total activity, which simplifies handling, disposal, and licensing requirements.

Consequently, this mode has dominated radionuclide wear measurements since the early 1970s, when several groups working with diesel engines independently developed SLA marker-based techniques (Ref 10, 11, 12, 13, 14). The first application of the marker approach in the United States was to measure missile nose cone ablation for the U.S. Air Force (Ref 15). A comparison of marker and tracer configurations is shown in Fig. 1. Both tracers and markers require corrections for natural radioactive decay, and possibly for background "noise."

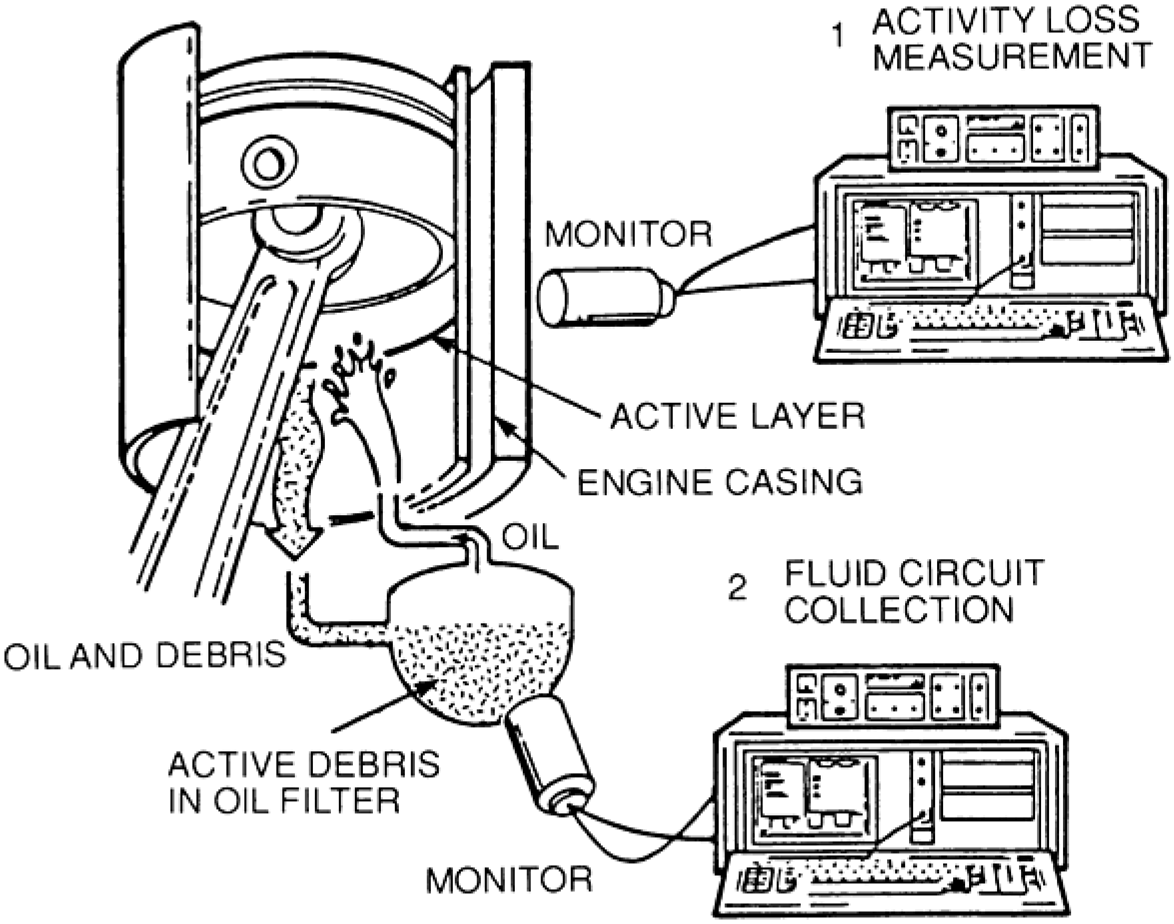


Fig. 1 Tracer and marker schemes commonly used for radionuclide wear measurements

In marker monitoring, a part or sample to be monitored is first exposed to a high-energy particle beam, which transforms atoms in a layer near the surface to specific radionuclides, about 1 atom per billion. Because total beam fluence is small, material properties do not change, and activity produced does not present any handling hazard.

Gamma Detectors. Both tracers and markers depend on gamma radiometry to deduce material amounts from detected gamma interactions. For marker monitoring, a gamma-ray detector is installed near the wearing part, typically a sodium iodide scintillator type (thallium-activated), shown in Fig. 2, to allow room-temperature spectroscopy. For tracers, a similar type of detector can be placed either at an oil filter or sump or simply near the recirculated oil line. Details of detector operation and optimizing performance are presented later in this article.

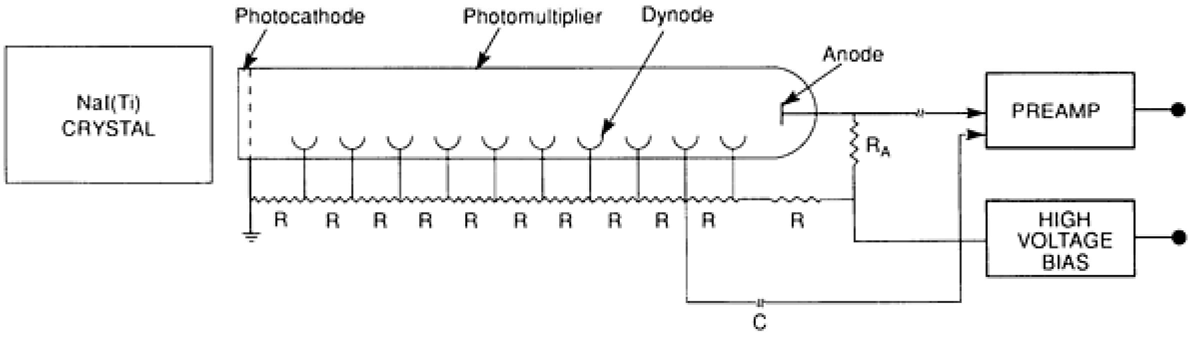


Fig. 2 Gamma-ray spectrometry based on sodium iodide scintillator that converts energy from gamma ray into visible light pulse

Activation Methods. The most commonly used methods for creating radionuclide tracers and markers (Fig. 3) are based on inducing nuclear reactions in the test piece itself. Either neutrons or energetic ions are allowed to interact, at the subatomic level, with a fraction of host matrix material. Lighter particles, such as electrons, photons, or mesons, are too unlikely to interact to be practical, and their range is too great to create a shallow layer.

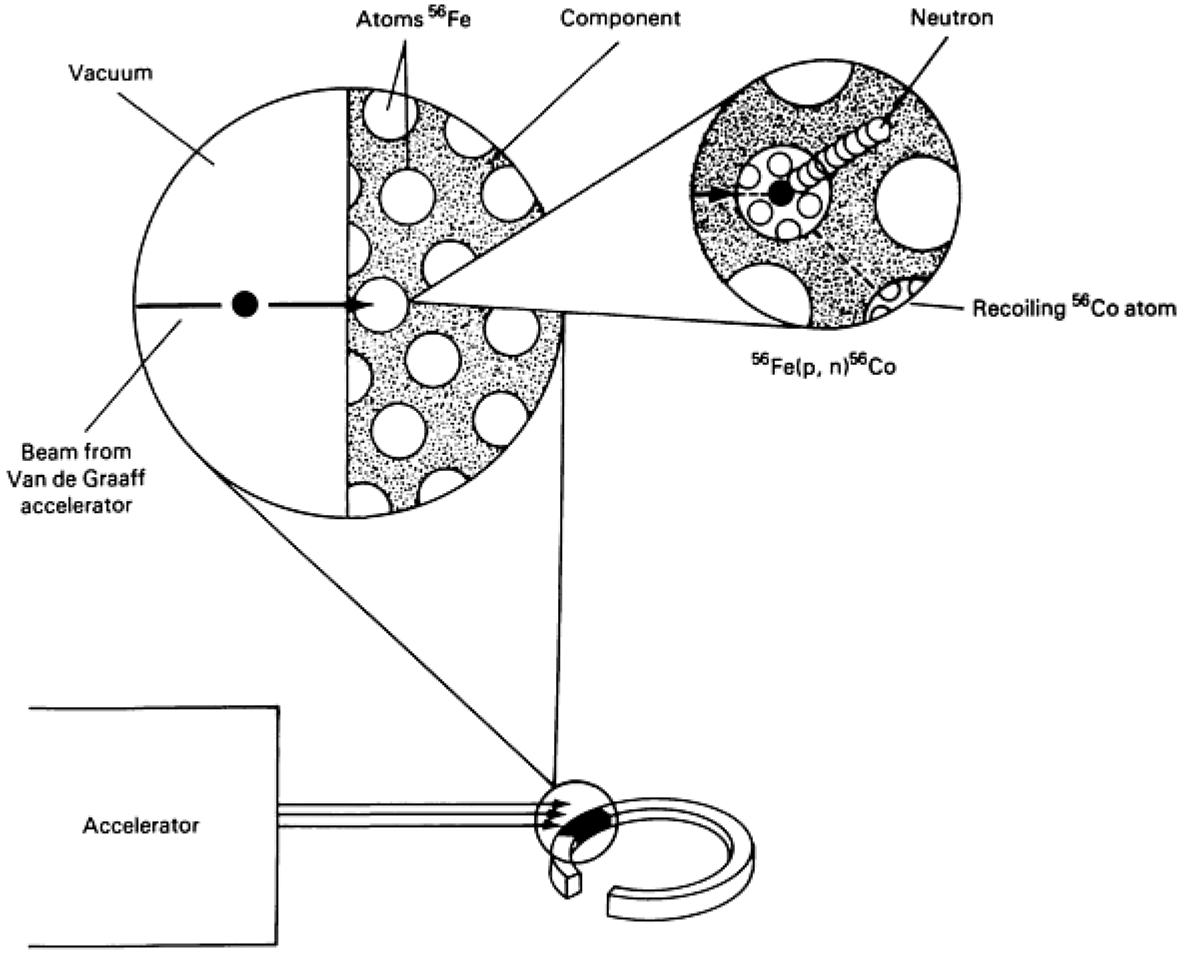


Fig. 3 Methods for creating radionuclides

The advantage of such in-place generation, compared to plating, coating, or implanting methods, is that the basic material properties of the host material are not changed. This allows valid comparisons between materials or coatings.

When the host material is too difficult to activate directly, either mechanical implanting or electrochemical plating can be selected for limited requirements, such as system characterization or lubricant evaluation. However, even for these restricted test objectives, care must be taken to minimize impact on the wear properties of the material itself, which could invalidate test indications. Radionuclides in the implants also typically come from either neutron or particle activation.

Thermal Neutrons. As previously mentioned, the first radionuclides in engine tests were produced by neutron activation. This is most commonly achieved by simply exposing the monitored part to thermal neutrons from a nuclear reactor core. Neutron activation can also be performed on-site using a portable neutron source, but these are rarely intense enough to create a usable activation. Care must be taken to avoid activating other materials in the vicinity. Small accelerators configured to produce neutrons can be used, but these are not portable, and they require substantial shielding when operating.

The main advantage of neutron activation is its relatively lower cost per microcurie of activity. However, there are many nuclides that can only be produced by accelerator reactions. For some materials, only accelerator-produced nuclides are suitable for monitoring, because of half-life or gamma energy. Conversely, nuclides produced by neutron absorption can also be generated by accelerated deuteron (H-2) bombardment.

Neutron activation has several major disadvantages for SLA purposes, compared to ion beams. Most notably, activity is necessarily distributed through the bulk material. This results in neutron activation being restricted almost exclusively to tracer-type monitoring. It is sometimes possible to partly shield or mask the activated component, but a shallow surface layer is not feasible unless the material to be activated is already in a thin coating.

Even with a thin coating, if the target constituent is not also present in the substrate to thoroughly confound the desired surface layer, then neutrons tend to activate other elements as well. Sometimes trace constituents can activate so much more readily than the one desired for monitoring that they create special handling problems. Consequently, neutron activation is difficult to configure without dealing with large amounts of unwanted activity, in addition to the relatively large amount required for tracer monitoring.

Energetic Ions. Modern particle accelerators can avoid these problems by carefully selecting the species and energy of the incident beam to stay below the reaction threshold for unwanted reactions. Often, depth of a single useful nuclide can be manipulated over a broad range without producing other gamma emitters, except for short-lived nuclides that will not interfere. Any allowed reaction only continues until the beam is below its unique threshold energy, which restricts activity to a shallow layer extending either several μm or sometimes up to several hundred μm from the surface.

This simple fact alone typically reduces the amount of activity needed by a factor of thousands in the ratio of the activated volume to the bulk volume. Switching from tracer to marker configurations can reduce the activity requirement further, depending on the attenuation involved. Thus, although cost per microcurie is much lower for neutron reactions, cost per monitored component is comparable for many materials.

Accelerator beams are typically narrow (pencil-sized) and easily restricted to small zones that are to be monitored using masks or collimation. Mask thickness only has to be great enough to exhaust the activation depth, although it is more common to completely absorb the beam range. To treat target areas larger than the beam spot (for example, the circumference of a piston ring), components can be either moved in the beam or rotated.

Among various types of ion accelerators, two are most commonly used for SLA (cyclotrons and tandem Van de Graaffs), because both operate in the energy range just above the threshold for common reactions. Cyclotrons typically reach energies well above this range, as well. Reaching the lower ranges that are best for SLA is difficult for many.

Even higher-energy machines, such as the synchrotron at Brookhaven National Laboratory, have been used to make centimeter-depth activations, but the need for this type of monitoring is relatively rare. High energy can also increase radiation damage and produce many unwanted nuclides through multiple-nucleon knock-out reactions. However, it can be advantageously used for recoil implantation, a process discussed in the next section.

For reactions involving low cross sections in durable materials, cyclotrons often have a cost advantage over less-energetic Van de Graaff accelerators, because they can generate large beam currents to shorten operating time and use extra beam energy to bring the beam out of vacuum. This keeps the irradiated sample from overheating in the higher beam current, and the convenience of bombardment without breaking vacuum is attractive. However, these advantages are obtained at the expense of degraded definition and spread in energy through straggling and beam dispersion.

Tandem Van de Graaff accelerators can also be configured to bring the beam out of vacuum, but, for most materials, this is not necessary (nor desirable) to create an economical bombardment. The advantage of lower beam currents delivered in vacuum is that the undegraded beam energy can be precisely reproduced with very little spread. This eliminates the need to calibrate the activity versus depth profile for each bombardment, the common practice with cyclotrons. In addition, many materials are sensitive to high currents and the resulting intense local heating, even with cooling at atmospheric pressure. Many materials that are too sensitive for any direct bombardment can still be activated using yet another activation technique, recoil implantation.

Recoil implantation can be used with both Van de Graaff and cyclotron accelerators to create ultrashallow activated layers (Ref 16, 17). In this process, a target foil is bombarded, instead of the part to be monitored, which is placed downstream of the foil (Fig. 4). Reaction products in the foil recoil with enough energy to escape and be implanted downstream. The result is a layer of activity without exposing the part directly to the beam.

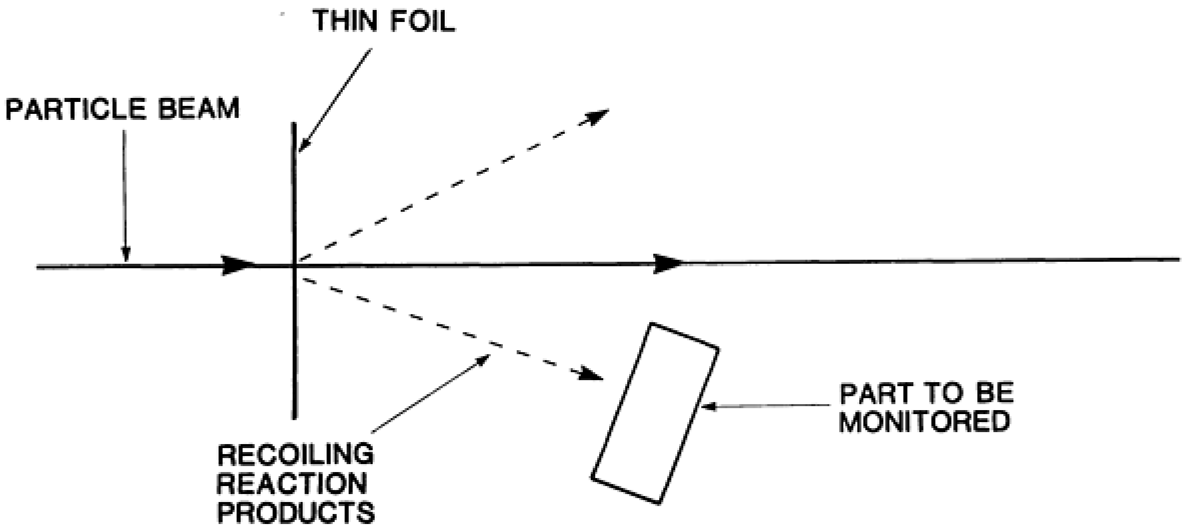


Fig. 4 Recoil implantation configuration

This is ideal for delicate materials, such as polymers, which would degrade from radiation damage, because of the direct beam. It is also an excellent approach for materials with low melting points. For shallow activation depths, this procedure is still usually performed in vacuum, because air or even helium would limit the range of large ions.

With the beam energies attainable in a tandem Van de Graaff accelerator, recoil-implant profile depths are typically under 1 μm (40 μin.). Such ultrashallow profiles are desirable in many precision measurements, such as magnetic recording media, or for detecting fretting damage.

Higher energies from cyclotrons and heavier projectiles allow reactions in which the beam ions fragment into lighter species. For example, the National Superconducting Cyclotron Laboratory at Michigan State University has created deeply implanted Be-7 profiles by nitrogen bombardment, in which the N-14 nucleus fragments into two Be-7 recoil atoms (Ref 18). The variety of reactions and depths available allows the activation of materials that are otherwise difficult to treat. However, high cost prevents deep recoil implantation from replacing direct bombardment in most materials.

Advantages of Surface Layer Activation. The extreme precision possible with SLA is its greatest single advantage, particularly for markers. It can greatly shorten test periods required to create measurable wear damage or reveal minute effects that are only accessible to its submicron resolution capability. With engine testing, the savings in fuel alone usually more than justifies the use of SLA techniques. Even in situations where other variables might interfere, such as temperature or pressure fluctuations, resolution of each increment of wear is still so fine that wear rates (relative measurements) are still more accurate than readings obtained by more conventional means.

In addition, both marker and tracer readings are made noninvasively, without interfering with any aspect of system operation. Readings are made *in situ,* without opening the system and possibly changing clearances, valve seating, or

vibration resonances. This is particularly important for lubricated systems, such as engines, which experience break-in wear whenever the system is opened.

Although static off-line readings are possible (and recommended for confirmation), SLA is typically configured to produce on-line monitoring in near real-time. This allows rapid feedback for either evaluating test conditions or preventing damage. Lubricants or other conditions can be modified on-the-fly, allowing the engineer to intelligently change test protocols, as needed, to investigate new phenomena.

The nature of the activation process makes SLA ideal for inaccessible locations, hostile environments, and parts with difficult shapes. Gamma-ray emission is not affected by environmental factors, and the signal is transmitted to a detector outside the system without wires, fiber optics, or connections of any kind.

Virtually any industrial metal and most other materials (ceramics, alloys, and many polymers) can be activated for wear monitoring. Nature was particularly generous with ferrous alloys, because four different nuclides can be generated from iron alone.

Simultaneous readings from several locations are possible through either one of two possible methods. Gamma-energy resolution, in which two distinct nuclides are used and monitored by one detector, is one method. Off-line analysis distinguishes gamma rays from each nuclide by their characteristic energies. This is ideal for simultaneously monitoring mating surfaces or for multiple tracers.

In the second approach (only possible with the marker method), several detectors are each collimated and shielded to view a single activated zone. Either multiple analysis systems are used to allow concurrent readings, or fewer devices are rotated between detectors for sequential readings. A switching system can be used to effect rapid changes.

Considerations in Planning SLA Testing

Choosing SLA for monitoring surface loss should be based on several key requirements. First, either marker or tracer methods are only valid for situations in which wear debris is actively removed from the measurement area. This can occur either through normal operation, or mating surfaces can be separated to determine whether material has transferred from one face to another.

Requirements for debris removal can be stretched somewhat, either by artificially interrupting a test to remove debris (for example, in fretting tests) or collimating the detector so that it only views the activated area, rather than an immediately adjacent zone. This approach has been used in sealed bearings, for example, but it is much more difficult than normal flowing configurations.

In addition, tracers require that wear debris be collected from the working fluid. For marker monitoring, the debris can be completely lost or exhausted, for example, from a gas turbine engine. In some systems, such as engines with recirculating lubricant, both methods are feasible and can be used simultaneously.

Other critical planning questions involve whether the material is suitable for activation at the required depth, whether a reliable monitoring configuration is possible, and whether the test information generated will, in fact, adequately answer the desired experimental questions. These issues are addressed in the section "Details for Executing SLA Measurements," which provides both a basis for evaluating SLA for a particular application and detailed procedures for conducting SLA monitoring.

Depth of the activated marker is an important parameter in the trade-off between measurement precision and test duration. It is controlled by beam energy and angle of incidence, once the reaction and, therefore, beam species, are selected. In general, limits to precision for individual wear increments are proportional to the total depth, usually by about 1%. Thus, a depth no greater than 100 μm (4 mils) is required for precision finer than 1 μm (40 μin.).

On the other hand, surface loss can only be monitored for as long as the marker persists. Thus, duration of a test is limited in proportion to marker depth, and a compromise must be made in test planning between precision and duration requirements. Sometimes, cost can play a role in making this compromise, because beam-time costs tend to increase as depth decreases.

Depths that are easily accessible to SLA range from 15 μm (0.6 mil) to hundreds of μm, depending on beam energy, which is often limited by the type of accelerator. Activation depths outside this range are certainly feasible, but they require alternate techniques, such as recoil implantation, or larger machines, which are either expensive or rarely accessible.

Typical Applications

One of the earliest applications of accelerator activation was associated with machine tooling, particularly drill bits (Ref 19, 20, 21). Parallel developments in measuring ablative wear in cannons occurred in this same period (Ref 22, 23, 24), as well as refinements in calibration and detector configuration for tracers in engines (Ref 25, 26, 27). Since these early tests, many successful applications have been made in industrial fields, including nuclear and fossil power generation, chemical processing, piping, medicine (prosthetic bone replacement), and aerospace propulsion, and research has been conducted in the fields of metallurgy, corrosion, and tribology (Ref 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54).

SLA markers are currently applied to a variety of industrial wear problems (including diesel engine parts such as bearings, piston rings, valves) and occasional large-scale (centimeter) erosion problems, such as in pipelines, reactor or furnace walls, boiler tubes, and other heavy industry components. Extreme precision of this technique is ideal for evaluating minute differences in wear rates, such as those produced by variations in critical operating conditions. Typical system parameters that can be correlated with these changes in wear rate include temperature, vibration, load (Ref 55), and, particularly, type of lubricant or lubricant conditions (Ref 56, 57). Coatings or surface treatments with very similar performance capabilities can also be rapidly compared (Ref 58).

For diagnostic condition monitoring, the convenience of on-line readings is frequently more significant than accuracy. Consequently, tracers are often used in industrial systems where debris can be caught in a filter for later examination in the laboratory. Of course, the tracer approach is only suitable for systems with an enclosed working fluid, such as diesel engines or steam turbines, and therefore could not be considered for gas turbine or jet propulsion applications, except possibly in bearings or a drive train. Tracers have been applied, with minimal success, to wear in magnetic disk drives (Ref 59). Ultrashallow markers have also been used for magnetic recording media (Ref 60).

Even in systems with a captured fluid, tracer monitoring can be less preferable, because it includes stochastic effects of mixing, settling, and chemical interactions, which can either delay or distort readings. In contrast, the only requirement for the marker approach is that debris be thoroughly removed from the immediate vicinity of the wearing surface so that it no longer contributes to the detected signal. Other considerations include whether the detector can withstand installation near the activated component (if not, tracers must be used), rapidity of measurements needed, and whether the user is willing to work with the higher levels of activity needed for tracers.

Laboratory Wear Characterization. Probably one of the most representative and consistently used applications of SLA is in lubricant testing. A good example is a continuing evaluation of additives for motor oil conducted at the Exxon Research and Engineering Company (Ref 56), based on the activation of overhead cam lobes in a four-cylinder 2.3 L engine. In tests lasting several hours, repeated marker readings track lubricated wear. An example plot of readings produced in this type of test is shown in Fig. 5. Note that there is some scatter in the data, which emphasizes that gamma-ray counting is an inherently statistical type of measurement. However, with a series of readings such as these, it is easy to determine the slope, or wear rate, for each operating condition.

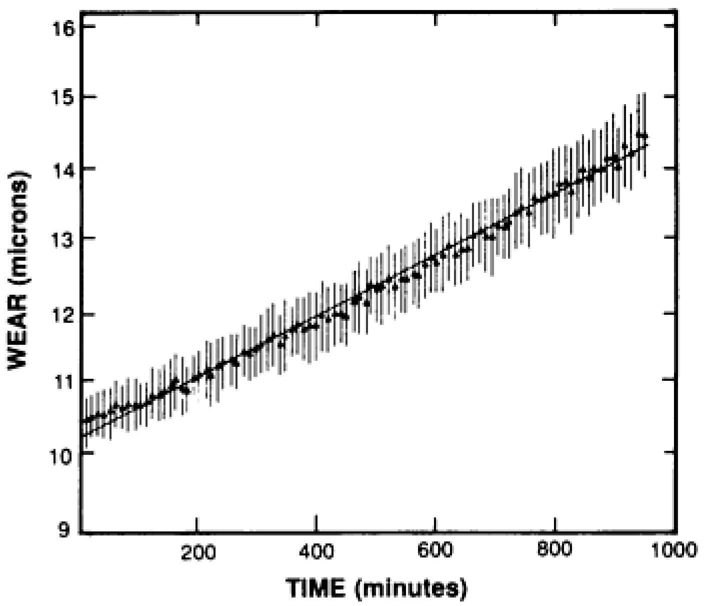


Fig. 5 Typical wear history. Source: Ref 56

Diagnostic Condition Monitoring. Field tests in which SLA was used to monitor heavy industrial systems have been effective, most frequently in fossil-fuel power plant components. For example, a configuration for monitoring a steam turbine stop/bypass valve at a power-generating station was recently tested by the Electric Power Research Institute (Ref 26). At the Stuart plant of Dayton Power and Light, the stop-valve bypass carries the entire steam flow to the turbine during start-ups. Oxide and magnetite exfoliation in upstream pipes and boiler tubes cause serious erosion of the valve, necessitating its annual replacement long before it is fully expended, which makes this valve an excellent candidate for monitoring.

Two activated valve disks were installed side-by-side on the same turbine inlet and monitored during a series of start-up cycles. One had a skirt made of conventional Stellite, and the other had a Stellite skirt with a boron diffusion treatment to increase erosion resistance. Scintillation counters monitored the activated regions of the valve disks through the valve body, approximately 300 mm (12 in.) of steel, plus insulation. Results are shown in Fig. 6.

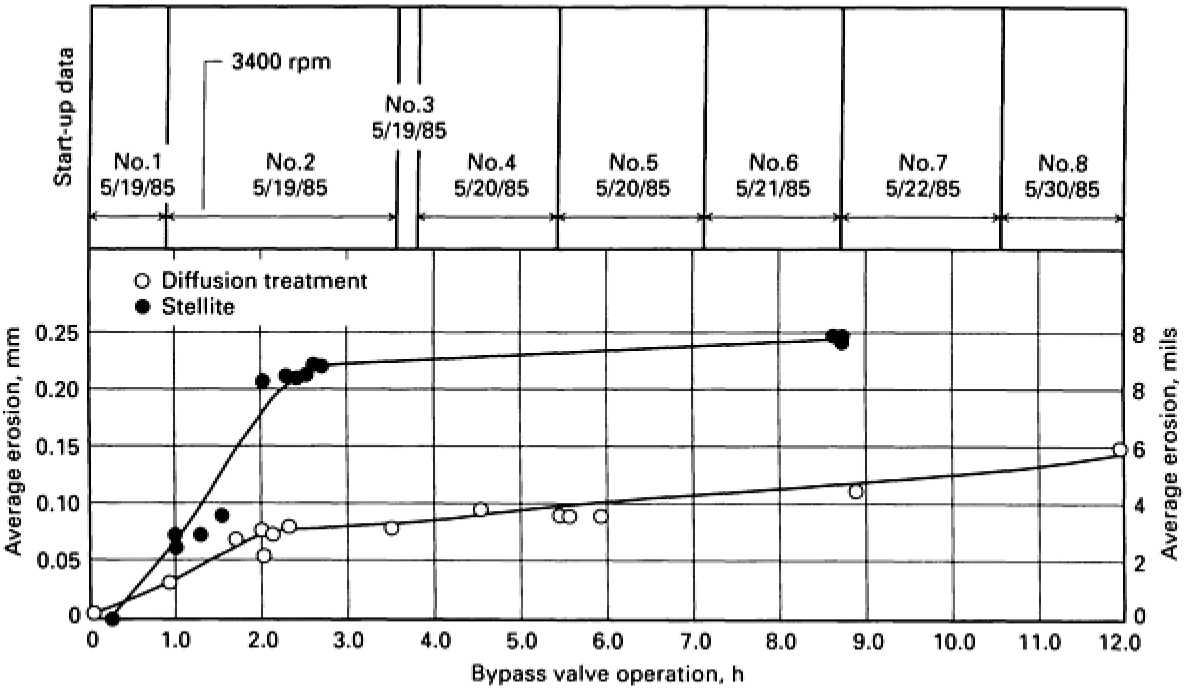


Fig. 6 Erosion of valve disks, which indicates ability of SLA to measure lower initial wear rate and eventual disappearance of boron diffusion coating

This test demonstrated that erosion resistance of the diffusion-treated valve was superior to that of the untreated valve during the first two start-up cycles. However, because the treated layer had eroded away by the end of the second start-up, the subsequent erosion rates of the two disks were essentially the same. Real-time monitoring provided a rapid and accurate comparison of the performance of the valve disks.

Tracer Monitoring. In heavy industry applications, measuring activity that accumulates in erosion debris is preferable to monitoring the part itself. For example, at another station (Ref 26), SLA tracers have been used to monitor high-pressure turbine blade erosion, because the thick casing makes direct measurement of blade activity impractical. Researchers sampled debris from the condensate water using a conventional felt bag filter, which was examined by gamma spectrometry at a laboratory away from the power plant. Multiple isotopes were used to independently detect erosion at different locations.

Engine testing has been the most common application for radionuclide tracers, as it is for markers. Tracers can have an economic benefit, in that they can achieve a degree of accuracy, even when based on less-costly activations in a nuclear reactor. However, because reactor activation is a bulk process, the entire component is uniformly affected, with the result that much more activity is needed than that amount required for the same accuracy with a thin layer. Often, the costs of handling, storage, and disposal after the test can outweigh the savings in the activation cost. Consequently, most tracer work today relies on cyclotron bombardment.

Tracers are still commonly used in U.S.-based companies (Ref 47), as well as in many other countries, as evidenced in recent literature from Germany (Ref 61, 62, 63, 64, 65), the United Kingdom (Ref 66, 67), Japan (Ref 68), France (Ref 69), India (Ref 70), Poland (Ref 71), and Ireland (Ref 72). A good example is a recent series of tests described in Ref 73, which includes a description of calibrations and procedures for measurements based on reactor activation. Such neutron activations can still be used to advantage in calibrating a cyclotron-based measurement, because it is easier to uniformly activate a powder using the reactor approach, than to polish a cyclotron-bombarded solid and collect the residue. However, a correction for detector efficiency can be required if the reactor-produced nuclide emits a different energy gamma ray from the cyclotron product.

Aerospace. In gas turbine engines, blades are often shrouded at their tips for aerodynamic control and structural support. A short segment of the shroud is attached to each blade, and these segments are separated by a small gap (Fig. 7). Clearances in these gaps are observed to grow during engine operation, either because of wear or deformation of the

shroud segments. Enlarged clearances allow increased vibrational amplitudes that lead to fatigue and eventual blade failure. Consequently, gaps are regularly inspected to ensure that total clearance remains within specified limits.

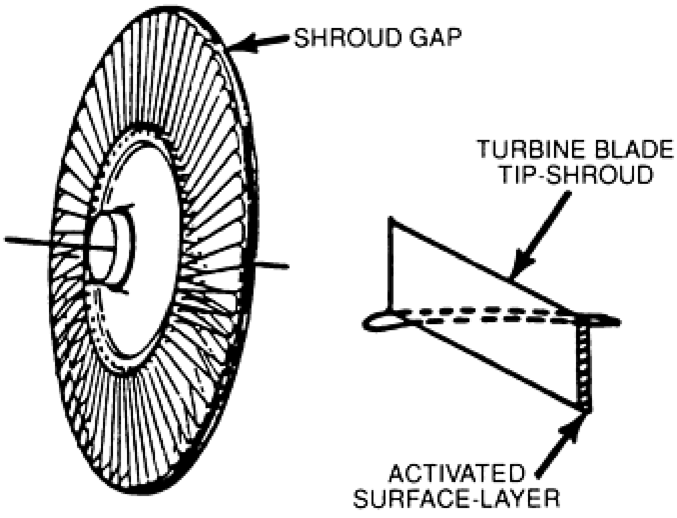


Fig. 7 Common arrangement of turbine engine blade shroud gaps. Source: Ref 55

Because inspection and repair are expensive processes, causes of gap growth and possible prevention were investigated using SLA markers (Ref 55). Two distinct nuclides, ^{60}Co and ^{54}Mn, were used to separately monitor gaps that were initially smaller or larger, respectively, than average. Two different turbine wheel configurations were tested, one "tight" with a total shroud gap of 0.41 mm (16 mils), and one "loose" with a total gap of 2.95 mm (116 mils).

On-line readings were able to characterize wear rates as a function of engine speed. These were found to be significant only at the highest speeds, 92 to 96%. A number of other significant results were produced. First, although the gap size increase was greater for the tight wheel, SLA indicated much greater wear for the loose wheel. This suggests that most of the initial increase in gap size for a new wheel is due to other processes, such as plastic deformation or turning of the blades, and not wear. Off-line readings for individual blades indicated that wear was a function of total gap size for the wheel, but was independent of individual gaps.

Another important indication from the SLA readings was the presence of a previously undetected vibrational mode. This was evident in anomalous increases in detected counting rate at a certain engine speed and believed to be related to opening and closing of shroud gaps during torsional motion. The motion was later linked to a resonance that could be suppressed by design changes.

Details for Executing SLA Measurements

Activation Fundamentals. Critical properties of gamma-emitting radionuclides, which determine their suitability for wear monitoring, are their decay half-line and gamma-ray energy, which must be sufficiently energetic to penetrate surrounding material and allow detection from the outside. Rate of decay must be sufficiently slow to allow a reliable signal to persist throughout the planned measurement interval. For quick tests, half-lives of a few days are sufficient. For long-term monitoring, reliable measurements can typically be made for up to six half-lives.

By carefully choosing the species of incident particle, bean energy, and angle of incidence, a suitable radionuclide can be produced in a controlled pattern. Lists of nuclides, their radiations, and reactions producing them are available in several well-known tables (Ref 74, 75). Some of the more commonly used species are summarized in Table 1. The reactions most commonly used to produce these are (p, n), (d, alpha), and (alpha, d), because these have low thresholds and large cross sections. In this nomenclature, the first symbol in the parentheses is the incident beam species, and the second is the reaction product. Letters stand for proton, neutron, deuteron, and alpha particle, respectively.

Table 1 Some useful SLA radionuclides from common metals and their production methods

Material	Accelerator beam	Major half-life radionuclide ($T_{1/2}$)
Cr	Alpha	^{54}Mn 300d
Cr	Protons	^{52}Mn 5.6d
Cu	Protons	^{65}Zn 244d
Fe	Alpha	^{58}Co 71d
Fe	Protons	^{56}Co 78d
Fe	Deuterons	^{54}Mn 300d
Ni	Deuterons	^{60}Co 5Y
Ti	Deuterons	^{46}Sc 84d
Ti	Protons	^{48}V 16d

Maximum depth of activity can be estimated by integrating the basic Bethe formula for ion stopping power (Ref 76). Low-energy corrections for recombination, adiabatic interactions, and (Sternheimer) density effects can be safely neglected, because interest is limited to range above the reaction threshold, which is typically at a relatively high energy level of several MeV.

An alternate approach is to look up ranges in available range tables (Ref 77). The range below the reaction threshold is subtracted from the range at incident energy to produce the depth of activity generated. Energy loss or range calculations can even be used to predict the activity versus depth calibration, if reaction cross sections are known at all involved energies. Cross sections are particularly important near the reaction threshold.

One problem with these calculations and range table estimates is that wear measurements are rarely performed in pure elements, and calculations are sensitive to average atomic number and density. Because exact composition and density of industrial metals can vary substantially, calculated profile calibrations are not reliable, even if reaction cross sections are available. Consequently, a laboratory calibration procedure should be completed for each combination of material and beam conditions.

Laboratory Marker Calibrations. To convert the decrease in marker intensity to a precise measurement of wear, the relation between the amount of material removed and the fraction of radiation intensity that remains must be known. This activity-versus-depth profile is typically determined in a series of calibration measurements by activating an identical part or sample and then carefully simulating wear by controlled polishing in the laboratory.

Care must be taken in mounting the specimen in a jig to ensure uniform surface removal from the activated area. Additionally, the grit size andtype of polishing fluid, if any, should be selected to match the wear or erosion process being simulated. Once begun, it is more important to maintain surface roughness than to actually match the surface finish on a worn part. Changing grit size or polishing conditions can cause an anomalous jump in the calibration curve.

Activity remaining is measured on an optical bench using a reproducible geometry for each increment of material and a technique appropriate to the depth involved. For deep profiles (more than 50 μm, or 2 mils), a precision micrometer is sufficiently accurate. A deep profile can also be evaluated by exposing a stack of foils made from the same material as the wearing part and then detecting the activity in each foil.

For shallower activation depths, an actual part can be exposed and then polished to simulate wear. To produce the necessary measurement resolution, either an electronic diamond-stylus profilometer or optical technique is required. For ultrashallow activations, both uniformity of surface removal and measurement of wear depth can become problems that require special care.

For profiles under 1 μm (40 μin.) deep, a diamond-tipped indenter (for example, in a microhardness testing system), is best. This approach is also useful in somewhat deeper profiles as a check against other measurement methods. Because the microscopic indentation is typically exhausted after 1 to 2 μm (40 or 80 μin.), several overlapped indents must be made to span the entire depth. Typical calibration microindents are shown in Fig. 8.

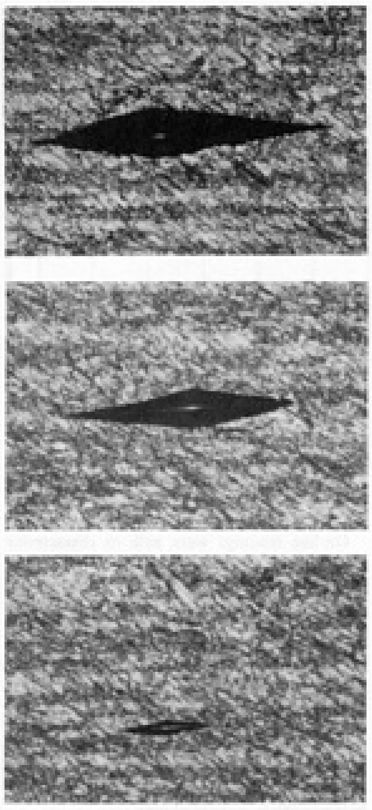


Fig. 8 Diamond-shaped (Knoop tip) indents before and after polishing. Longest indent is about 31 μm (1.2 mils) long and 1 μm (40 μin.) deep.

Once the distribution of activity remaining as a function of depth removed is known, the original part can be installed for actual wear monitoring or testing. A representative calibration curve is shown in Fig. 9.

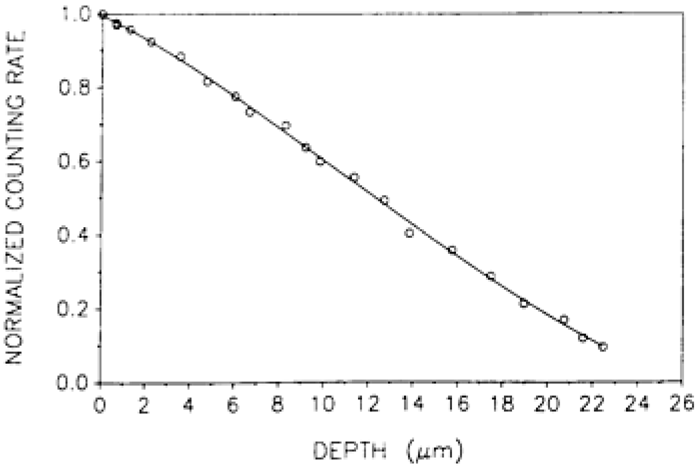


Fig. 9 Typical activity versus depth profile calibration curve. Linearity implies relatively uniform concentration of activity.

Tracer Calibrations. Tracer monitoring also requires calibrations, but not of activity versus depth, which is usually assumed to be uniform. Instead, a correlation between total activity, distributed in an unfiltered operating test system, and

detected counting rate must be made. If the detector is stationed at a filter, then activity in the filter must be correlated with detected counting rate.

Both filtered and unfiltered calibrations can be achieved by adding a known amount of activated debris (or simulated wear debris) to the lubricant or filter. A good check on this type of calibration is to add activated material in small increments to confirm a linear relation between counting rate and total added activity. Linear regression analysis on incremental readings also smooths out small measurement errors from each sample. An example of this type of calibration for a small bearing test rig, rather than a full engine, is shown in Fig. 10.

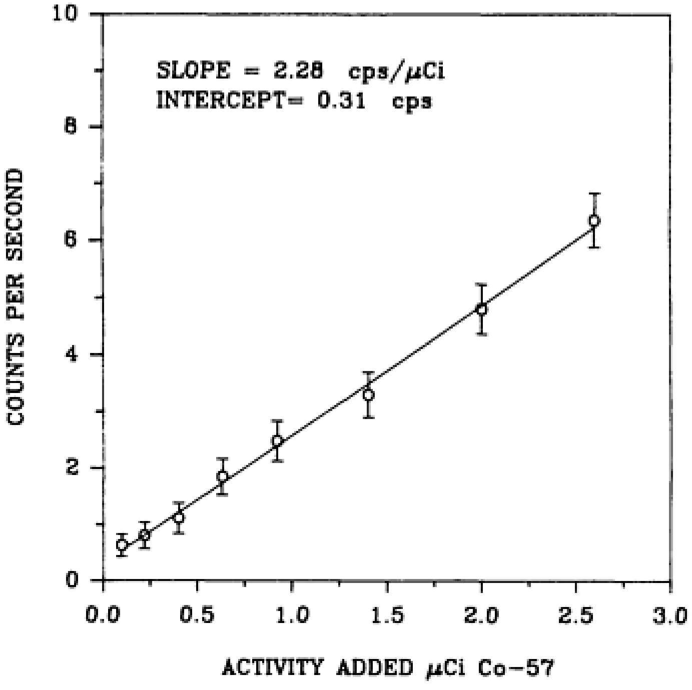


Fig. 10 Counting rate measured by a 50 × 50 mm (2 × 2 in.) detector as a function of activity dispersed in a lubricated bearing test rig with circulated oil

Attenuation and Detection of Gamma Rays. Interactions of gamma rays with matter critically determine SLA performance, both by creating scintillator responses and by attenuation of signal. The same mechanisms are responsible for both detection and attenuation: primarily Compton scattering, but also photoelectric absorption and electron-positron pair production.

Only a few nuclides with gamma energies that are sufficient to produce positrons also have long enough half-lives to be suitable for SLA monitoring. Photoelectric absorption is only prominent for low energies or very high atomic numbers. Consequently, interactions essential for both detection and attenuation are dominated by Compton scattering in most SLA configurations.

It is often useful, in planning an SLA measurement, to estimate the detected counting rate based on distance from source to detector, source strength, gamma energy, thickness of attenuator, and size of the detector scintillator. This estimate is important for optimizing the signal from a monitored component and minimizing interference from background nuclides. Choice of nuclide, source strength, and detector size may all need revision if the counting rate is too low or too high.

Measured intensity from a point source (small compared to the distance, *r*, to the detector) is proportional to $1/r^2$. This means they should generally be as close as practical to ensure a strong signal. Signal strengths for a 76 × 76 mm (3 × 3 in.) detector can be more than adequate several feet away from quite modest sources, but when attenuation is added, the detector may require closer positioning to avoid signal-to-noise problems in brief readings.

The $1/r^2$ dependence occurs because the detected signal is proportional to the effective solid angle (surface area subtended on a sphere of unit radius) intercepted by the cross-sectional area of the detector. However, this is exactly $1/r^2$ only when

detector and source are both small, compared to the distance. In fact, sensitivity to changes in range is enhanced for detectors that are moderately large, compared to *r*. The size of this increase can be $1/r^3$ or better over a selected range. In contrast, at extremely close range, measured intensity becomes insensitive to changes in *r*.

Attenuation in intervening material in many systems has a stronger effect on detected signal than range. Intensity falls exponentially on passage through matter with a strong dependency on energy, density, and atomic number (*Z*). Because attenuation depends on different processes, effects of atomic number vary with gamma-ray energy. Also, at low energies where the dependence is approximately Z^3, discontinuities occur because of x-ray shells.

The complexity of these dependencies is concealed in the relatively simple attenuation formula:

$$I = I_0 \, e^{-\mu \rho x} \qquad \textbf{(Eq 1)}$$

where I is the detected radiation flux, I_0 is the source flux, μ is the mass attenuation coefficient, ρ is the mass density, and x is the linear range through the material in question. The mass attenuation coefficient, μ, can be represented as a sum of attenuation coefficients for each of the separate interaction processes:

$$\mu_{(\text{total})} = \mu_{(\text{photo})} + \mu_{(\text{Compton})} + \mu_{(\text{pair})} \qquad \textbf{(Eq 2)}$$

As previously discussed, Rayleigh and other very low energy photon interactions are negligible above a few keV. Therefore, the corresponding additional terms are missing. Because $\mu_{(\text{total})}$ contains the complicated dependencies on materials, values are most easily obtained from readily available tables (Ref 78, 79, 80). Representative plots for steel and the detector material are provided in Ref 81.

Presence of mass density, ρ, in the exponent implies that the denser the material, the more rapidly the signal is attenuated. Consequently, the detector should be positioned to avoid intervening materials, particularly dense or large atomic number materials, which can block the signal or create a shadow effect.

Partial shadowing can be especially troublesome if the source moves at all with respect to the shadowing object during measurements. Materials such as steel or, especially, lead and tungsten can also be used to collimate or shield sources for storage and shipment.

In addition to solid angle effects, the issue of appropriate detector size is complicated for high-energy gamma rays by thickness of the scintillator, which can greatly affect efficiency. Lower-energy interactions are unaffected by extra scintillator material if they are certain to occur near the front surface.

Detector Configuration. Following calibration, the activated part can be installed with a gamma-ray detector nearby. Sodium iodide scintillator (activated by thallium) is usually preferred, because of its relatively high detection efficiency and stability in light output when operating at room temperature. Even large crystals are relatively inexpensive, which is important for working with small or heavily attenuated sources with penetrating hard gamma rays.

Energy resolution in NaI(T1) is at least a factor of two better than similar materials, such as plastics or bismuth germanate. Most SLA applications are in low-background environments and involve only one or two nuclides. Consequently, high-resolution provided by cryogenic semiconductor detectors (Ge or GeLi) is not needed. Bismuth germanate does have advantages, when cold, in detection efficiency (particularly important for high-energy gamma rays) and resistance to thermal shock.

Alkali halide scintillator detectors were invented in 1948 at Stanford University (Ref 82), and several good reviews are available (Ref 83, 84). The material is relatively dense (3.67 g/cm^3), and it has a high effective atomic number that is due to iodine (Z = 53) in the compound. This results in a high electron density, which is an important parameter in determining overall registration efficiency.

Characteristics of Sodium Iodide Spectrometry. When gamma rays either Compton-scatter or are absorbed through the photoelectric effect in the scintillator material, recoiling or ejected electrons interact to produce a cascade of secondary electrons. These also scatter before coming to rest, until all of the energy from the original interaction is

deposited. The result is to excite electrons that remain electrostatically bound to the hole produced in the valence band, forming an electron-hole pair, or "exciton."

These excitons can wander through the lattice until they encounter the thallium activator (about 0.1 to 0.3% doping), an impurity that creates a high probability of secondary excitation, followed by radiative decay. The overall effect is to cause a relatively fixed portion of the original energy deposited to transform to light by fluorescence of the impurity. Thallium was selected for wavelengths that match the photocathode response curve and to which NaI is transparent. These are broadly peaked near 413 nm and range from 300 to over 500 nm, well into the visible wavelengths.

Decay time for this light pulse (to 1/e of original intensity) is about 230 ns, which is typically larger than the photomultiplier transit time, but smaller than the amplifier shaping time (usually about 1 μs). Much faster electronic designs are available, but are not needed for the very low counting rates typical of SLA. Constant fraction discrimination or time-of-pulse-peak determination using differentiation and zero-crossing circuits can even circumvent the phototube and scintillator time restrictions to achieve subnano-second resolution.

The basis for SLA is that total fluorescence intensity is proportional (with statistical variations from the number of photons involved) to the excitation produced. This is, in turn, proportional to energy deposited by the original gamma ray. The major function of the multichannel analyzer, then, is to sort these pulses by height, producing a spectrum plotting intensity (or counts) as a function of detected gamma energy. The process is therefore called pulse-height analysis. A schematic diagram of this process is shown in Fig. 11.

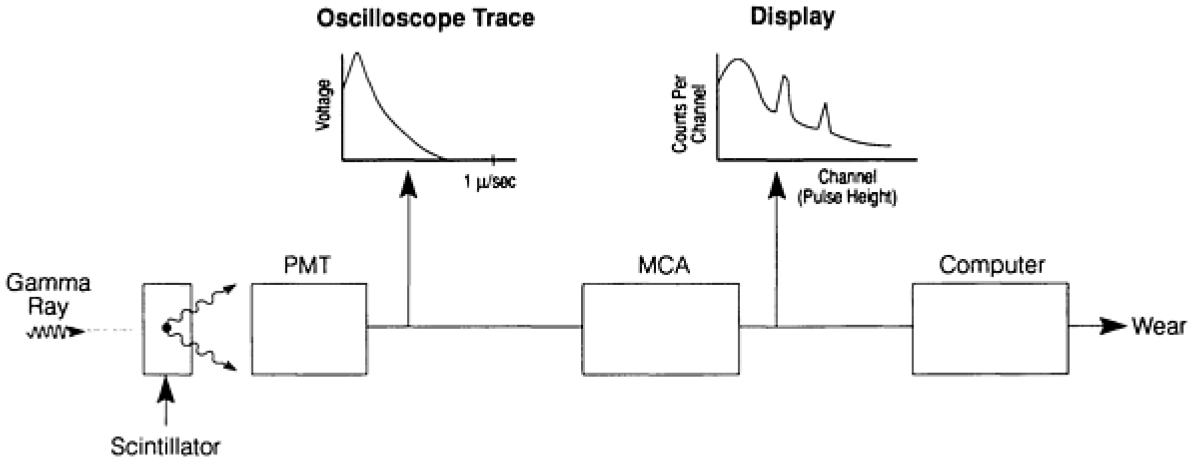


Fig. 11 Components in pulse-height analysis

Spectrum Details. The resulting spectrum has a number of features, some of which are useful, and some of which either constitute an interfering background or combine to form a noise continuum. The most important feature for SLA, and spectrometry in general, is the so-called photoelectric peak produced by complete absorption of the original gamma-ray energy in the detector. This occurs most frequently through photoelectric absorption or through secondary absorption of a Compton photon that also scattered in the scintillator. The object for SLA is to resolve a photopeak from others and from background contributions. Many techniques for achieving this resolution and interpreting the results are described in available literature (Ref 85, 86, 87).

If a Compton photon escapes after scattering, the partial energy deposited registers below the full-energy photopeak. Because this energy and cross section for scattering are angle dependent, Compton scattering combined with photon escape creates a broad plateau. The upper energy boundary is referred to as the "Compton edge" formed when the gamma ray scatters backward 180° from incidence.

A complementary Compton peak often appears below the Compton edge, because of photons absorbed after backscattering outside the scintillator. The positions of the Compton edge and peak can be calculated using the Compton equation:

$$E' = E/[1 + (1 - \cos\theta)\, E/mc^2] \qquad \textbf{(Eq 3)}$$

where E and E' are the incident and scattered photon energies, θ is the scattering angle, and mc^2 is the rest mass energy of the electron, 0.511 MeV.

The recoiling electron energy and, therefore, the registration energy given by $E - E'$ have a maximum value for backward scattering, where $\theta = 180°$. Consequently, the Compton edge occurs at:

$$e_{max} = \frac{2E^2}{mc^2 + 2E} \qquad \textbf{(Eq 4)}$$

and the Compton peak occurs at:

$$E_{max} = \frac{E}{1 + 2E/mc^2} \qquad \textbf{(Eq 5)}$$

Photons detected after multiple scattering form a low-energy continuum, which merges with electronic noise because of random thermo-electric emission in the photomultiplier and amplifier circuits. Representative spectra are shown in Fig. 12 and 13. For very weak signals, noise that is due to externally generated background photons can be reduced somewhat by shielding the detector.

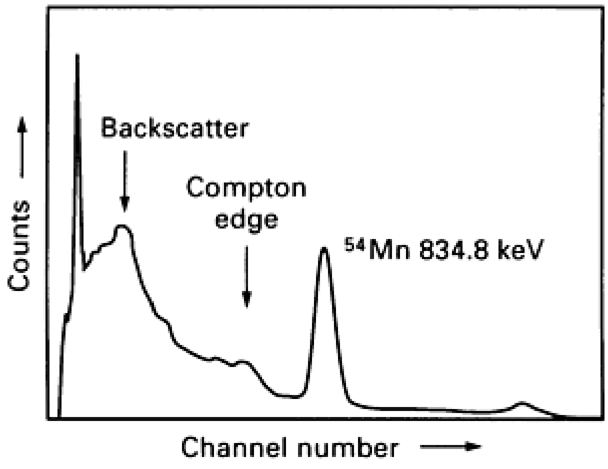


Fig. 12 Typical gamma spectrum for ^{54}Mn

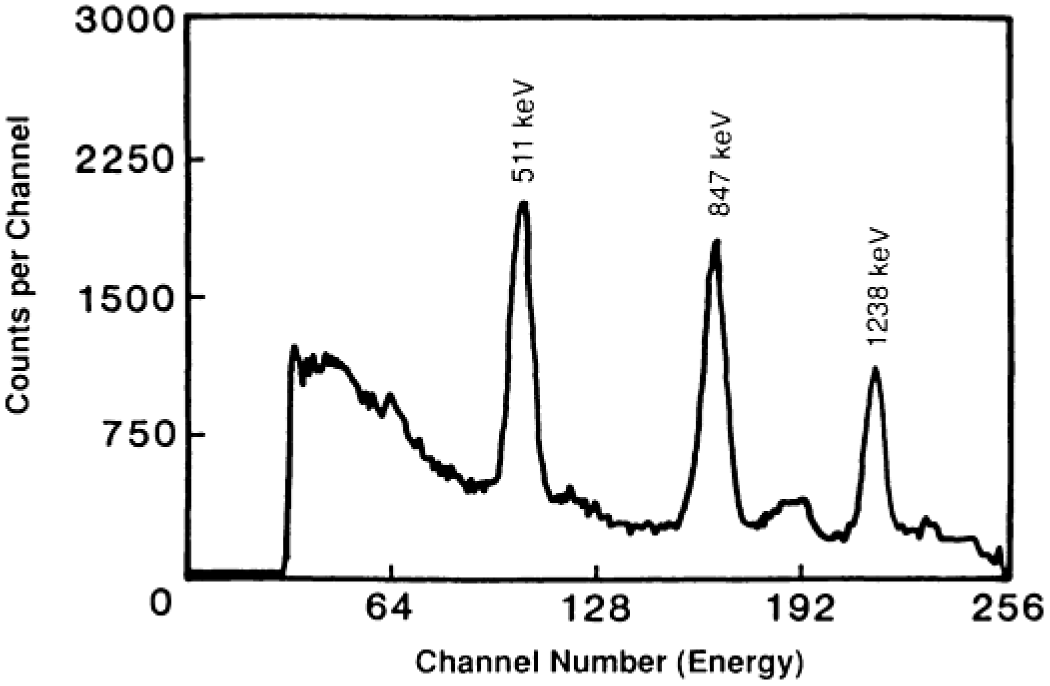


Fig. 13 Typical gamma spectrum for ^{56}Co with multiple peaks indicated

Statistical Considerations. Data reduction to separate peaks from background is critical to ensure reliable SLA measurements. In most situations, spectra are relatively simple in that they contain contributions from only one or two nuclides. The only difficulty is that the signal strength is often low, and prolonged counting may be needed to ensure minimal statistical variation.

At least 10,000 counts are needed under an isolated photopeak to ensure a 1%, one uncertainty. When the signal is comparable to background "noise," shielding or coincidence arrangements can be used to improve the separation of signal counting rate from background. Because this is effectively a subtraction, uncertainties add in quadrature. Therefore, many more counts are needed under the peak to ensure precision.

Source strength, attenuation, and detector efficiency together determine measured counting rate under the peak which, in turn, determines length of each measurement interval to produce a particular level of scatter in the data. For marker measurements, this relation of interval and variability is virtually fixed, once the detector and source are installed. Intervals can only be shortened during operation by sacrificing some precision. Consequently, some planning is needed to ensure a reasonable counting rate. A larger source, larger detector, or better mounting position may be needed.

For markers, the best counting rate occurs at the start of monitoring. As testing continues, both surface loss and natural decay reduce the counting rate. Thus, either measurement intervals must increase or precision will decrease with testing. An artificial limit of 90% of the profile depth is well advised, because similar limits to accuracy apply in producing the profile calibration.

For tracers, the original signal is zero, or very close to it. When wear begins, the signal typically jumps dramatically to a nonzero value, and counting statistics continue to improve as wear increases. This gives tracers and advantage in detecting onset of wear, because it is generally easier to detect a doubling or tripling of a small counting rate, than to detect a 1% drop in a larger marker counting rate.

Detector Mounting. Because the way the detector is mounted can create other adverse effects, some planning is again needed to ensure an effective mounting. Dominant concerns are to:

- Maximize counting rate while avoiding cross talk (from other sources or debris accumulations)
- Ensure a fixed geometry relative to source and attenuator

- Minimize temperature changes
- Avoid hazards to detector integrity (vibration, shock, and rapid temperature changes)

As previously discussed, counting rate can be increased by reducing either the source-to-detector range or attenuation, or both. However, this is generally subject to constraints on threats to the detector. In proximity to engines or power systems, the detector often must be wrapped in a blanket of thermal insulation or water cooled (Fig. 14). The mounting frame must be heavy enough to prevent large-amplitude vibrations or motion.

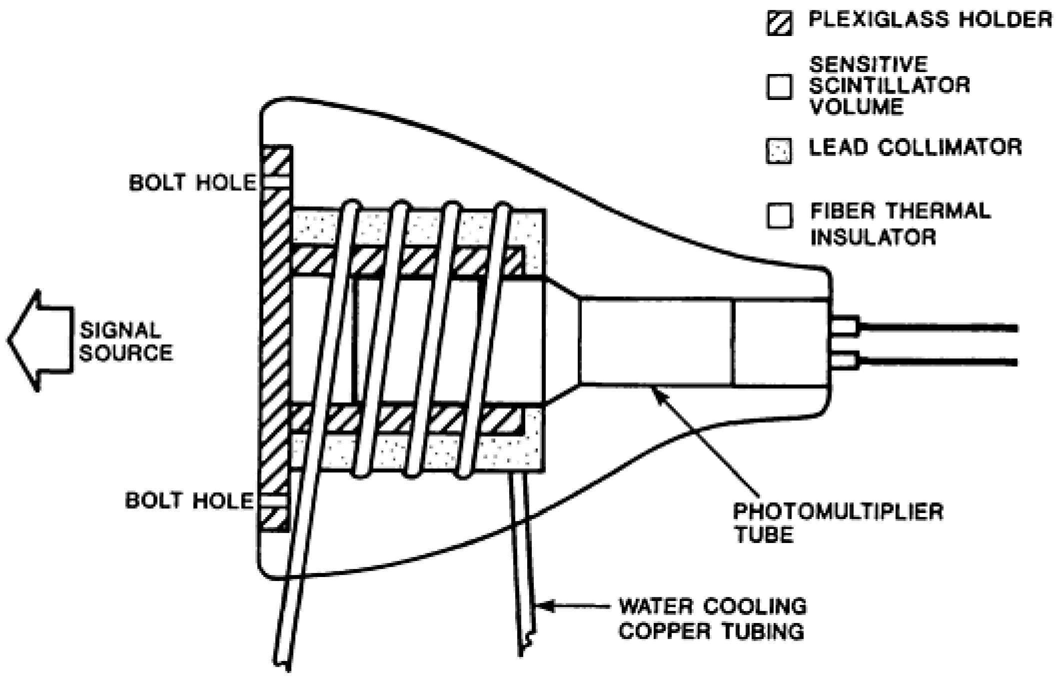


Fig. 14 Commonly used detector insulation and cooling scheme

Another concern is that wear debris containing the active nuclide must be positively removed from the area of the detector to produce a valid measurement. Generally, this is guaranteed by early planning of the program, which considers the action of lubricants, or fluids, in the system.

Sometimes additional shielding is needed to prevent debris in oil lines, traps, and filters from continuing to affect the detector. If other sources are monitored in the same system or if radiation is in the test environment (thorium is often used in engines to make refractory components), then the detector may require shielding all the way around, creating a telescopic tube, or collimation.

Temperature of the detector is another possible concern, particularly if operation of the system is intermittent so that heat flow frequently changes. In addition to the need to protect the detector from thermal shock that is due to rapid temperature changes (greater than about 7 °C, or 12 °F, per hour for larger NaI crystals), it is possible that photo-conversion efficiency and photomultiplier tube gain can both be affected by temperature.

Fortunately, efficiency changes are relatively minor effects near room temperature for NaI(Tl), as shown in Fig. 15. This is an important advantage for this type of detector. Operation at higher temperatures is also possible, but stability becomes more critical. Because the phototube produces its own heat, starting with a cold detector that is below 10 °C (50 °F) can cause thermal shock to the scintillator or damage to the dynodes.

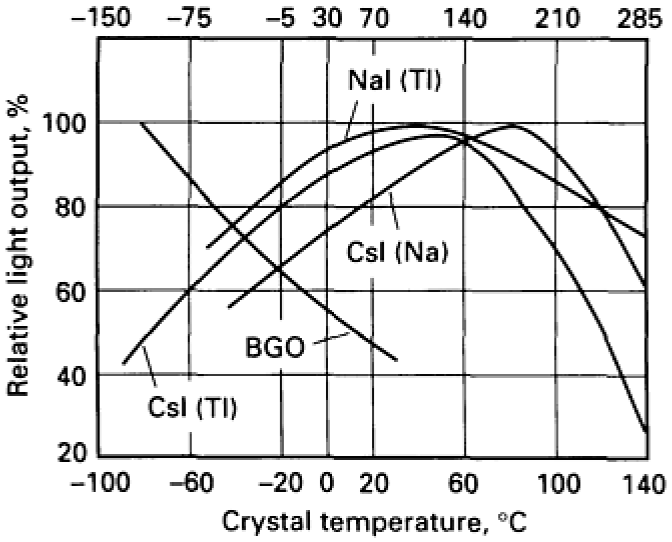


Fig. 15 Scintillator efficiency as function of temperature for several detector materials

Temperature-based changes in photomultiplier gain are primarily an annoyance, causing the photoelectric peak to move, but the area under the peak should be constant. There can be other causes of peak drift, as well. All can be corrected in analysis or by adjusting the amplifier gain to compensate.

Attenuation changes that are due to expansion of the mechanical system being monitored are only significant if temperature varies over a very large range during measurements and heavy attenuation is involved. The size of the effect depends on details of the system, but for less than 100 mm (4 in.) of steel, changes of less than 275 °C (500 °F) can usually be neglected. When thermal expansion effects are significant, initial no-wear reference readings should not start until the system is warmed to near operating temperatures. In some systems, measurable wear or erosion occurs during this warm-up. Therefore, readings must be extrapolated back to determine a warmed, but unworn, reference counting rate.

A more insidious, but potentially more problematic effect of system warm-up is due to movement of heavily attenuating components relative to the detector. These can cast a shadow on the detector, which moves as the system changes temperature. Thermal expansion causes minute effects, compared to those of a moving shadow. Consequently, care must be taken to avoid geometries with mobile shadows, even if a reduced average counting rate results.

Future Trends

SLA tracers and markers have been demonstrated to be ideal for unobtrusive measurement of μm-scale wear in operating systems, with many successful applications to engine components, such as pistons, cylinders, rings, liners, bearings, cams, and gears. Large industrial systems have also been effectively monitored, including steam turbine components, chemical reactor systems, piping, valves, orifice plates, fittings, and nozzles. Surface layer activation can also be used to study changes in composition or structure in off-line analysis to optimize engineered materials.

The precision offered by this type of technique can produce significant savings by reducing the test time required to produce measurable surface loss or to distinguish materials that are similar in durability. Parameters that influence surface loss can be varied rapidly to optimize conditions or to determine effects of related processes such as chemical (corrosive) reactions.

Consequently, the future of surface layer activation is quite promising. Because technical requirements have been thoroughly explored in the last two decades, new applications have become almost routine. In particular, techniques for producing reliable measurements from minimal sources, such as shallow markers, efficient detectors, collimation, and analysis methods, have made surface layer activation attractive to many smaller testing facilities that would not consider the reactor-nuclide-tracer methods common in the 1960s. Most modern marker applications do not even require licensing.

There are also high-temperature surface loss problems in industry and propulsion that benefit from this type of precision measurement. Because detectors operate outside the monitored system, it is relatively simple to protect them from harsh environments and heat. Industrial systems that operate at more moderate temperatures, but in corrosive or erosive environments, have already benefited from SLA capabilities.

The ability to precisely measure durability on-line, or to study migration of elemental components off-line, will be critical to development of advanced materials and coatings. SLA measurement offers a convenient and extremely precise way to meet modern requirements for *in situ* monitoring.

Eventual on-board wear monitoring in both ground (Ref 88) and flight (Ref 89) vehicles is also expected to be valuable, if not absolutely necessary, in systems that are designed to operate near critical design limits. Surface layer activation may eventually be required for routine diagnostic monitoring of critical engine components. Deeper profiles could be used for such long-term applications, and readings might be incorporated into the instrument displays to warn of potential component failure.

References

1. P.L. Pinotti, D.E. Hull, E.J. McLaughlin, Application of Radioactive Tracers to Improve Automotive Fuels, Lubricants, and Engines, *SAE Quarterly Transaction of the National Passenger Car Body and Production Meeting,* Mar 1949
2. A.K. Hannum, "Radioactive Tracers in Engine Research," Thompson Products Engineering Bulletin No. 1, Vol 2, 1956, p 1
3. H. Halliwell, Radiotracers Reveal Engine Wear During Detergent Oil Filtration Study, *SAE Trans.,* Vol 68, 1960
4. J.J. Gumbleton, Piston Ring and Cylinder Wear Measurements Illustrate the Potential and Limitations of the Radioactive Technique, *SAE Trans.,* Vol 70, 1962
5. J.E. McClelland and S.M. Billett, "Filter Life versus Engine Wear," SAE 650866, Fuels and Lubricants (Tulsa), Nov 1965
6. T.A. Tennyson and C.K. Parker, Jr., "Locomotive Radioactive Ring Studies of Fuel, Lubricant, and Operating Variables," SAE 700892, National Combined Fuels and Lubricants and Transportation (Philadelphia), Nov 1970
7. J. Volz, "Erstellung optimierter Einlaufprogramme von Dieselmotoren (Determination of Optimized Break-In Schedules for Diesel Engines)", dissertation, Gesellschaft für Kernforschung mbH, Karlsruhe Laboratorium für Isotopmentechnik, KFK-2432, Mar 1977
8. A. Gerve, *Kerntechnik,* Vol 14, 1972, p 204
9. M.B. Treuhaft and J.O. Storment, "Investigation of Radioactive Tracer Filter Performance Tests," Southwest Research Institute, final report, SwRI Project 11-9189, Jan 1980
10. A.J. Armini, "SPI-WEAR, A Real Time Wear Monitoring System," TR-75-07, Spire Corporation, 1975
11. T.W. Conlon, Thin Layer Activation by Accelerated Ions--Application to Measurement of Industrial Wear, *Wear,* Vol 29, 1974, p 69-80
12. A. Gerve and G. Schatz, *Proceedings of the 7th International Conference on Cyclotrons and Their Applications,* Birkauser, Basel, 1975, p 496
13. B. Herkert, *Proceedings of the 7th International Conference on Cyclotrons and Their Applications,* Birkauser, Basel, 1975, p 514
14. K.V. Ettinger, J.H. Fremlin, and N.A. Askouri, *Proceedings of the 7th International Conference on Cyclotrons and Their Applications,* Birkauser, Basel, 1975, p 518
15. A.J. Armini and S.N. Bunker, *Inst. Soc. Am.,* Vol 15, 1976, p 22
16. J. Asher, T.W. Conlon, J.H. Shea, and D.J. Parker, *Nucl. Instrum. Meth.,* Vol 178, 1980, p 293
17. T.W. Conlon and D.J. Parker, *Nucl. Inst. Meth.,* Vol 177, 1980 p 613
18. M.L. Mallory, R.M. Ronningen, W.C. McHarris, B. Sherrill, Y.X. Dardenne, and H.J. Schock, MSUCL-658, Michigan State University, Sept 1988

19. B. Ivković and A. Misković, Measurement of Spiral Drill Wear with New Radioactive Method, in *Proceedings of the 3rd I.C.P.R.* (Amherst), 1975

20. B. Ivković Experimental Investigation in Production Engineering by Radioactive Method, in *Proceedings of the 4th I.C.P.R.* (Tokyo), 1977, p 885-898

21. V.I. Postnikov, Method of Thin Layer Activation in Industry, *Moscow Atomizdat* (in Russian), 1975

22. A. Niiler and S.E. Caldwell, Measurement of Wear from Steel Using the Radioisotope ^{56}Co, *Nucl. Instrum. Meth.,* Vol 138, 1976, p 179

23. A. Niiler and R. Birkmire, *Nucl. Instrum. Meth.,* Vol 149, 1978, p 301

24. A. Niiler and R. Birkmire, Use of the ^{56}Fe(p,n)^{56}Co Reaction in Erosive Wear Measurements, *IEEE Trans. Nucl. Sci.,* Vol NS-26 (No. 1), Feb 1979, p 1815

25. W. Lausch, *Kerntechnik,* Vol 18, 1976, p 475

26. G.W. Jones, A.J. Armini, and N.C. Schoen, "Selective Radioactive Tracers for Engine Wear Diagnosis and Filter Evaluation," SAE 780972, International Fuels and Lubricants (Toronto), Nov 1978

27. W. Kaiser, Gesellschaft für Kernforschung mbH, Karlsruhe, KFK-Bericht 1568, Feb 1982

28. P. Loges, C. Blatchley, and D. Grandjean, Tracer Monitoring of Steam Turbine Seals: An Application of SLA, T.R. Shives, Ed., *43rd Mechanical Failures Prevention Group,* National Bureau of Standards, Oct 1988

29. C. Blatchley and P. Loges, "Monitoring Solid Particle Erosion in Turbines by Debris Analysis," ASME Joint Power Generation Conference (Boston), Oct 1990

30. C. Blatchley and P. Sioshansi, "On-Line Characterization of Wear in Mechanical Systems: Studies Using SLA," American Society of Lubrication Engineers, May 1987

31. "Wear Measurement by Surface Layer Activation," CS-5163, Electric Power Research Institute, May 1987

32. C. Blatchley, Real-Time, On-Line Durability Measurements in Advanced Materials Using Surface Layer Activation, *Proceedings of NDT&E of Advanced Materials and Composites Conference,* G. Matzkanin, Ed., NTIAC, Southwest Research Institute, USAF Academy, Aug 1986

33. P. Sioshansi and C. Blatchley, "Surface Layer Activation: A Novel Technique for Measurement of Wear-Corrosion-Erosion," 3rd Berkeley Conference on Corrosion-Erosion-Wear at Elevated Temperatures (Berkeley), Jan 1986

34. C. Blatchley *et al.,* "Boiler Circulation Pump Wear Monitoring: A Novel Technique Based on Surface Layer Activation," ASME 85-JPGC-Pwr-48, Joint Power Generation Conference (Madison), 1985

35. D Duncan *et al.,* "Diagnostic Monitoring of Erosion in Utility Steam Turbines," ASME 85-JPGC-Pwr-57, Joint Power Generation Conference (Madison), 1985

36. P. Sioshansi and C. Blatchley, "Surface Layer Activation Technique for Monitoring and In-Situ Wear Measurement of Gas Turbine Components," AIAA/SAE/ASME 20th Joint Propulsion Conference Cincinnati), June 1984

37. C. Blatchley and P. Sioshansi, "SLA: Precise Wear Measurements for Cogeneration Systems," ASME Internal Combustion Engine Technical Conference (Kansas City), Oct 1987

38. P. Sioshansi and C. Blatchley, "Real Time Wear Measurement of Coal Slurry Erosion Using Accelerator Induced Radionuclides," International Symposium on Slurry Flows (Anaheim), ASME, Dec 1986

39. C. Blatchley, P. Sioshansi, and R. Leyse, "Water Wall Erosion Precision Monitoring: An Application of Accelerator-Induced Activation," EPRI Conference on Power Plant Performance Monitoring and System Dispatch Improvement (Washington), Nov 1986

40. C. Blatchley and P. Sioshansi, "On-Line Wear Monitoring in Bearings and Gears Using Radionuclide Markers," T.R. Shives, Ed., *41st Mechanical Failures Prevention Group,* Naval Air Test Center, Oct 1986

41. C. Blatchley and P. Sioshansi, "Millisecond Bearing Wear: Surface Layer Activation in a Gas Turbine Engine" and "Erosion Monitoring in Oil Field Pipeline Equipment by Deep Layer Activation," *Proceedings of the 40th Mechanical Failures Prevention Group,* T.R. Shives, Ed., National Bureau of Standards, April 1985 (also published by Cambridge University Press, Cambridge, UK, 1987)

42. M. Yamamoto, "Utilization of Radiations in Tribology," SAE Japan 89-13, Symposium on Advanced

Tribology in Automobiles, 1990

43. P. Leterrible, M. Ducreux, M. Guerrand, G. Blondiaux, M. Valladon, and J.L. Debrun, Corrosion Studies in Reactor Secondary Cooling Systems, *Nucl. Instrum. Meth.,* B10/11, 1985, p 1054
44. D.H.J. Goodall, T.W. Conlon, C. Sofield, and G.M. McCracken, Thermonuclear Fusion Reactor Wall Erosion, *J. Nucl. Mat.,* Vol 76/77, 1978, p 492
45. J. Asher, R.F.A. Carney, T.W. Conlon, N.J.M. Wilkins, and R.D. Shaw, *Corros. Sci.,* Vol 23 (No. 4), 1984, p 411
46. T.W. Conlon, *Contemp. Phys.,* Vol 23 (No. 4), 1982, p 353
47. E.W. Schneider, D.H. Blossfeld, and M.A. Balnaves, "Effect of Speed and Power Output on Piston Ring Wear in a Diesel Engine," SAE 880672, Mar 1988; also in GMR 6033, General Motors Research Laboratories, Oct 1987
48. J.F. Wakenell, S.G. Fritz, and M.B. Treuhaft, "Diesel Fuel Specification and Locomotive Improvement Program: Eighth Research Phase Final Report," Southwest Research Institute Report 03-1542 and Association of American Railroads Report R-697, Dec 1987
49. R. Barron, "Engineering Model for Mechanical Wear," Report CMLD-CR-10-87, David Taylor Naval Ship Research and Development Center, June 1987
50. C. Blatchley an P. Sioshansi, SLA Technique for Monitoring and *In Situ* Wear Measurement of Turbine Components, *Jet Propulsion 3,* No. 3, May-June 1986, p 248-252
51. B. Ivković, Application of TLA to Measuring Tool Wear in Metal Cutting, *Tribol. Int.,* Vol 1, 1982
52. B. Jeanneau, Use of Thin Layer Activation Techniques to Assess Mechanical Parts Wear: A Survey of Industrial Experience in France, *IEEE Trans. Nucl. Sci.,* Vol NS-30 (No. 2), April 1983, p 1614
53. M. Bouchacourt, C. Marsigne, G. Blondiaux, and J.L. Debrun, Study of Corrosion-Erosion Phenomena in Nuclear Power Plants using Thin Layer Activation Method, *Bull. Am. Phys. Soc.,* Vol 33 (No. 8), 1988, p 1737
54. B. Natter, "Etude de l'Usure et de l'Ecaillage de Pieces d'un Reacteur d'Avion par la Technique d'Activation en Couches Superficielles (Study of Wear and Galling of Aviation Engine Components by the Technique of Surface Layer Activation)," unpublished Ph.D. dissertation, l'Universite Louis Pasteur de Strasbourg, CRN/PN 87-29, No. 321, Dec 1987
55. P.S. Kuo and C. Blatchley, "Turbine Blade Tip-Shroud Wear Characteristics Monitored by Surface Layer Activation," ASME 89-GT-300, Toronto, June 1989
56. H. Shaub and L.L. Wong, "Real Time Radioactive Marker Technique for Measuring Valve Train Wear," SAE 872156, International Fuels and Lubricants (Toronto), Nov 1987
57. S.G. Fritz and G.R. Cataldi, *In Situ* Piston Ring Wear Measurements in a Medium-Speed Diesel Engine, *Lubr. Eng.,* Vol 46 (No. 6), 1989, p 365
58. G. Wallace, L.H. Boulton, and D. Hodder, Corrosion Monitoring on a Large Steel Pressure Vessel by Thin-Layer Activation, *Corrosion,* Vol 45 (No. 12), 1989, p 1016
59. F. Talke, "Wear and Boundary Lubrication of Oxide Coated Magnetic Disks," NSF Tribology Program Workshop (Atlanta), Aug 1987, and private communication
60. Blatchley, "Monitoring Wear in Engine Components by SLA: Extension to Nanometer Accuracy," SAE Fuels and Lubricants Conference (Toronto), Nov 1987
61. C. Eifrig, Application of Thin-Layer Activation to Radiometric Wear Measurements, *Schmierungstechnik,* Vol 20 (No. 7), 1989, p 199-201
62. K. Eichorn, E. Richter, and H. Taubert, "Use of Thin-Layer Activation in Cavitation Studies," Technical Report ZFK-594, Zentralinst. Kernforsch., Rossendorf, Dresden, 1986, p 22-24
63. A. Kleinarahm and P. Fehsenfeld, Recent Developments in Wear Measurements with Thin Layer Activation Technique, *KFK-Nachr.,* Vol 21 (No. 1-2), 1989, p 22-26
64. Ch. Eifrig, K. Eichhorn, and P. Hammer, Thin Layer Activation for Wear Studies in Compressors, *Nucl. Instrum. Methods Phys. Res. B,* Vol B27 (No. 3), 1983, p 458-461
65. G. Essig and P. Fehsenfeld, Thin Layer Activation Technique and Wear Measurements in Mechanical

Engineering, *Proceedings of the 7th Conference on Nuclear Physics Methods in Materials Research,* (Vieweg, Braunschweig, FRG), 1980, p 70-81

66. R. Evans, Radioisotope Methods for Measuring Engine Wear: A Thin Layer Activation Method for the Measurement of Cam Follower Wear and Its Comparison with a Neutron Activation Method, *Wear,* Vol 64 (No. 2), 1980, p. 311-325
67. T.W. Conlon, Thin Layer Activation for Materials Analysis, *Ind. Lub. Tribol.,* Vol 34 (No. 1), 1982, p 20-25
68. T. Kosako and K. Nishimura, Wear Measurement at Depths of Several Tens of Micrometres on the Surface of Iron Using a Thin Layer Activation Method by 7 MeV Proton Beam, *Nucl. Instrum. Meth. Phys. Res. B,* Vol B40-41, 1980, p 587-590
69. A. Gallmann, B. Natter, and M.A. Molinari, Study of Wear and Galling in Aircraft Fuel Pump Drive Shafts and Gears Using the Surface Layer Activation Technique, *Nuc. Instrum. Meth. Phys. Res. B,* Vol B34, 1988, p 479-482
70. D.P. Chowdhury, J. Chowdhuri, A. Chakrabarti, B.B. Bhattacharjee, and S. Gangadharan, Study of Wear Between Piston Ring and Cylinder Housing of an Internal Combustion Engine by Thin Layer Activation Technique, *Indian J. Phys. A,* Vol 71A, 1988, p 688-690
71. W.H. Scharf, Thin Layer Activation for Wear and Corrosion Measurements--Present Status and Future, *J. Trace Microprobe Tech.,* Vol 5 (No. 4), 1988, p 269-299
72. S. Jetley, Choosing the Basic System in the Application of Thin Layer Activation Technique to Cutting Tools, *Wear,* Vol 80, 1982, p 125-127
73. G.R. Wendel, S.G. Fritz, and M.B. Treuhaft, Wear Analysis of Hydraulic Components Using Radioactive Wear Techniques, *Proceedings of the 43rd National Conference on Fluid Power* (Chicago), Oct 1988, p 217-225
74. C.M. Lederer and V.S. Shirley, *Table of Isotopes,* 7th ed., Wiley-Interscience, 1978
75. E. Browne and R.B. Firestone, *Table of Radioactive Isotopes,* Wiley-Interscience, 1986
76. M.S. Livingston and H.A. Bethe, *Rev. Mod. Phys.,* Vol 9, 1937, p 245
77. C.F. Williamson, J.P. Boujot, and J. Picard, "Tables of Range and Stopping Power of Chemical Elements for Charged Particles of Energy 0.05 to 500 MeV," Saclay Report CEA-R 3042, Department de Physique Nucleaire, July 1966
78. J.H. Hubbell, "Photon Cross Sections, Attenuation Coefficients, and Energy Absorption Coefficients from 10 keV to 100 GeV," NRDS-NSB 29, National Standard Reference Data System, National Bureau of Standards, 1969
79. W.J. Veigele, E.L. Briggs, B. Bracewell, and M. Donaldson, "X-ray Cross Section Compilation," KN-798-69-2(R), Kaman Nuclear Corporation, Oct 1969
80. F. Biggs and R. Lighthill, "Analytical Approximations for X-ray Cross Sections," SC-RR-66-452, Sandia Corporation, Feb 1967
81. R.D. Evans, Chapter 25, *The Atomic Nucleus,* McGraw-Hill, 1955
82. R. Hofstadter, Alkali Halide Scintillation Counters, *Phys. Rev.,* Vol 74, 1948, p 100
83. J.B. Birks, *The Theory and Practice of Scintillation Counting,* Pergamon Press, Oxford, 1964
84. R.L. Heath, R. Hofstadter, and E.B. Hughes, Inorganic Scintillators: A Review of Techniques and Applications, *Nucl. Instrum. Meth.,* Vol 162 (No. 1-3), Part II, 1979, p 431
85. J.H. Neiler and P.R. Bell, *Alpha-, Beta-, and Gamma-Ray Spectroscopy,* Vol 1, K. Siegbahn, North-Holland, 1965, p 245
86. L.G. Kanipe, S.K. Scale, and W.S. Liggett, "Least-Squares Resolution of Gamma-Ray Spectra in Environmental Monitoring," Report EPA-6001 7-77-089, U.S. Environmental Protection Agency, 1977
87. R.E. Wood and J.H. Palms, A Gamma-Ray Spectrum Analysis Technique for Low-Level Environmental Radionuclides, *IEEE Trans. Nucl. Sci.,* Vol NS-21 (No. 1), 1974, p 536
88. E.W. Schneider and D.H. Blossfeld, "Real-Time Measurement of Camshaft Wear in an Automotive Engine--A Radiometric Method," SAE 902085, Society of Automotive Engineers, 1990

89. P. Loges, C. Blatchley, E. Tobin, and P. Sioshansi, "SLA for Measuring Fretting Wear," Society of Tribologists and Lubrication Engineers Annual Meeting (Montreal), April 1991

Introduction to Laboratory Characterization Techniques

A.W. Ruff, National Institute of Standards and Technology

MANY WELL-ESTABLISHED TECHNIQUES for the measurement and characterization of bulk materials and their surfaces have been successfully applied in published tribological investigations. In addition, specialized techniques have been developed in many instances to provide information for unique questions that have arisen in such investigations. The following articles discuss the most commonly used characterization techniques along with selected special methods, and they provide both examples and references to successful applications. The purpose of these presentations is to give the reader a broad picture of the choices available among the laboratory characterization techniques, and also to provide brief descriptions of the fundamentals involved with each technique. The emphasis in the presentations is on the methods themselves, on equipment requirements, and on illustrative examples.

Most of the characterization techniques discussed in these articles are reasonably well developed (light microscopy, for example); however, their application to friction, lubrication, and wear investigations has introduced new and difficult challenges. The authors of the following articles are expert in the techniques described, and in most cases they have some specific experience with tribological applications. However, the range of issues associated with tribological technology is so wide that readers are cautioned to be alert to significant differences that may exist between their application and any particular example being discussed.

The techniques described in the following articles would customarily be applied in a laboratory setting to either laboratory specimens from testing or research programs, or to specimens intended for or removed from service, that is, actual components in tribological systems. In any case, it is essential that *all* parts of the tribological system be considered. There is ample discussion elsewhere in this Handbook of the complexity of friction, lubrication, and wear phenomena. Both the type and the significance of each process involved in the overall tribological phenomenon varies from case to case. The correct and useful application of information obtained from the applied characterization techniques depends on a thorough, comprehensive consideration of all aspects: the solid bodies, the environment, and any damage products from the operation or test.

This caution to the investigator cannot be overemphasized. As an example, consider the premature wear-out of a component in a lubricated mechanical system. Such a problem could be caused by periodic overloading due to transient excessive loads, or by overheating and degradation of the lubricant, or by corrosion effects resulting from process stream leakage, or many other reasons. Correct analysis of this problem can hardly be expected if information is not obtained on the loading history of the device, or on temperature excursions of the lubricating system, or on corrosive contaminants in the lubricant, to list only a few areas of concern. Failure in a tribological system is frequently the result of various factors acting in combination. For this reason, the investigator usually must apply a combination of characterization techniques and integrate the results in order to obtain a correct explanation.

The order of presentation of the articles on the individual techniques is somewhat arbitrary, and it is not meant to represent a recommended sequence for any particular investigation. Each author presents some discussion of the appropriateness of the technique, and the type of information that it yields, in order to guide the reader in making choices in this complex area.

Surface Texture

J.F. Song and T.V. Vorburger, National Institute of Standards and Technology

Introduction

TRIBOLOGISTS are concerned with what happens when two solid surfaces slide over each other, under either dry or lubricated conditions. They often measure and compare the surface texture and roughness parameters before the after the wear process, and at intermediate stages as well. Understanding the relationship between wear properties and surface texture can lead to the specification of optimized surface textures and manufacturing processes for various surface function needs.

The term "surface texture" refers to the fine irregularities (peaks and valleys) produced on a surface by the forming process. By convention, the texture comprises two components; roughness and waviness. Roughness consists of the finer irregularities characteristics of the process itself, such as the grit spacing of a grinding wheel or the feed of a single-point tool. Waviness consists of the more widely spaced irregularities that are often produced by vibration in the machining process. Usually, however, the terms "surface texture" and "roughness" are used interchangeably, because roughness is specified and measured much more often than waviness. The surface topography includes these surface texture components as well as any other irregularities, such as error of form.

There are two types of profiling techniques for surface texture, depending on whether the height measurements are made simply along a line (profile methods) or over an area (raster area methods). Both types of techniques develop quantitative knowledge of the surface peaks and valleys by point-by-point measurement with a high-resolution probe. Surface profiles are generally measured across the surface lay, the direction of uniaxial machining marks caused by most surface forming processes. The profile may be produced by a contacting stylus that traverses the surface, by a noncontacting probe, or by optical interferometric fringes. The stylus method has been widely used (Ref 1, 2, 3, 4, 5, 6, 7, 8, 9, 10) by engineers, opticists, and tribologists. Therefore, the stylus technique will be discussed in some detail in this article as an important example of surface profiling. Other techniques are discussed elsewhere (Ref 11, 12).

Surface Statistics

In general, surface textures are highly complex, because many surface finishing processes, such as polishing, grinding, and shot blasting, are statistical by nature. To characterize such surfaces, two types of statistical descriptors are used: parameters, such as root mean square (rms) roughness, which attempt to quantify some aspect of the surface statistics with a single number, and statistical functions, such as the power spectral density, which by their nature yield an array of information about the surface.

Surface Parameters

There are a great variety of surface parameters, many of which have been developed to characterize the function of engineering surfaces for particular applications. In fact, about 50 to 100 parameters have been defined for industrial use, and many of these appear in national standards as well. Nevertheless, surface parameters can generally be classified as height parameters, wavelength parameters, shape parameters, and combinations of these, known as hybrid parameters.

Height Parameters. The most common statistical descriptors of surface height are the roughness average, R_a, and the rms roughness (also called R_q). These are closely related and are given by the following formulas, shown in integral and digitized form:

$$R_a = \frac{1}{L}\int_0^L |y(x)|\,dx = \frac{1}{N}\sum_{i=1}^{N} |y_i| \quad \textbf{(Eq 1)}$$

$$R_q = \left[\frac{1}{L}\int_0^L y^2(x)dx\right]^{1/2} = \left[\frac{1}{N}\sum_{i=1}^{N} y_i^2\right]^{1/2} \quad \textbf{(Eq 2)}$$

where $y(x)$ is the surface profile, sampled by the set of N points y_i over the length, L, as shown in Fig. 1. The parameters R_a and R_q are useful estimators of the average heights and depths of surface profiles. The rms roughness is commonly specified for the surfaces of optical components. In general, the lower the rms roughness of an optical component, the less stray scattered light and thus the higher the quality of the component. Roughness average is used in the automotive and other metalworking industries to specify the surface finish of many types of components, ranging from cylinder bores to brake drums (Ref 13). In addition to these two averaging height parameters, an assortment of other height parameters have been defined for various applications, including several for characterizing peak-to-valley height (Ref 4, 14).

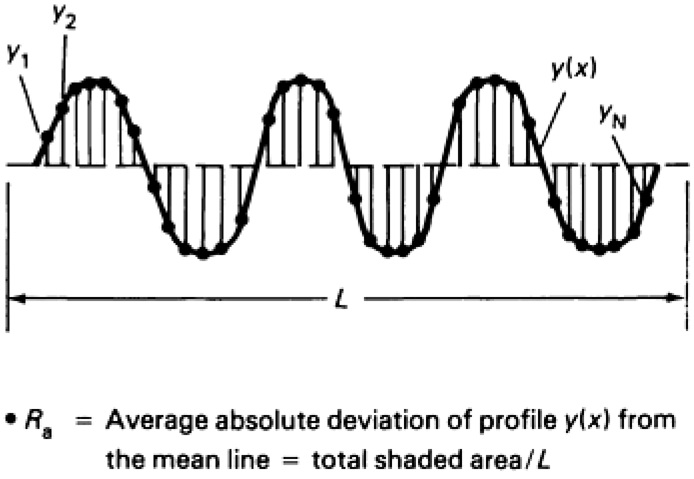


Fig. 1 Idealized stylus profile showing the mean line; the evaluation length, *L*; the digitized points, y_i; and a definition of the parameter R_a

Wavelength parameters are used to characterize the spacings of the peaks and valley of the surface. The spacings or wavelengths are often characteristics of the process that formed the surface, such as the shot size used for abrasive blasting, the grit size of a grinding wheel, or the feed of a tool. A typical wavelength parameter, recognized as standard by the International Organization for Standardization (ISO), is the mean peak spacing S_m (Ref 14), defined for a surface profile as the average spacing between two successive negative crossings of the mean line (see Fig. 2).

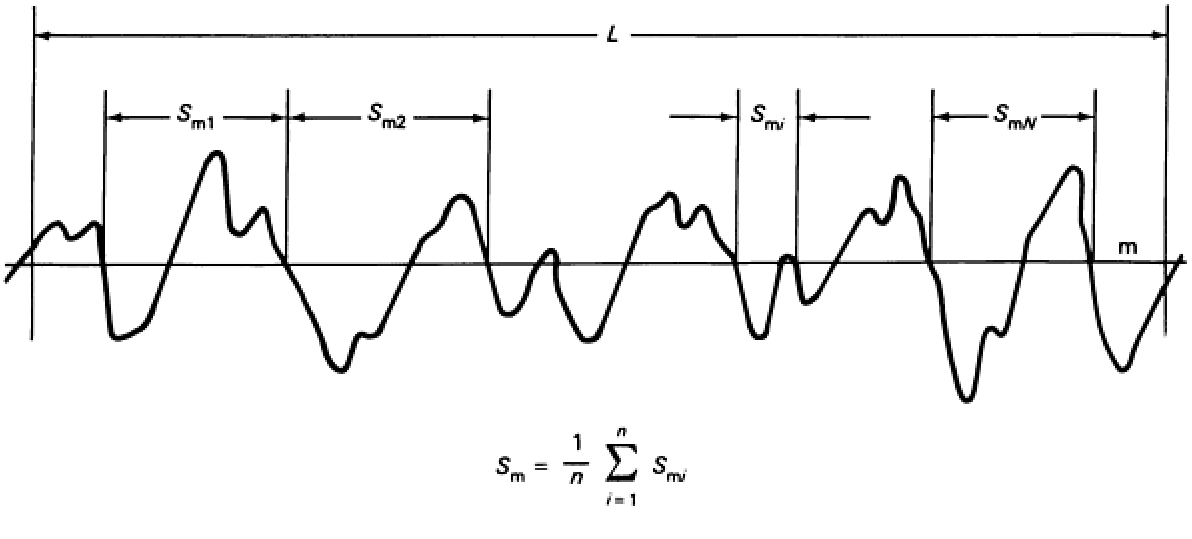


Fig. 2 Surface profile showing the ISO definition for the peak spacing parameter, S_m

Shape Parameters. The periodic profiles in Fig. 3 all have the same R_a and wavelength, but have different shapes and thus may perform differently in different applications. In particular, the profile in Fig. 3(b) represents a good load-bearing surface, and the profile in Fig. 3(a) represents a poor one. Because of its facets, the profile in Fig. 3(c) is best used as a diffraction grating, whereas one of the other surfaces may be most suitable for lubricated sliding. Shape parameters help to quantify the differences between these surfaces. The most important parameter, the skewness, is a measure of the symmetry of the profile about the mean line. It is defined as:

$$R_{sk} = \frac{1}{NR_q^3} \sum_{i=1}^{N} y_i^3 \quad \textbf{(Eq 3)}$$

According to this definition, the skewness of the profile in Fig. 3(a) is positive, whereas its opposite number (Fig. 3b) has negative skewness. The other two profiles (Fig. 3c and Fig. 3d) are symmetrical and have zero skewness.

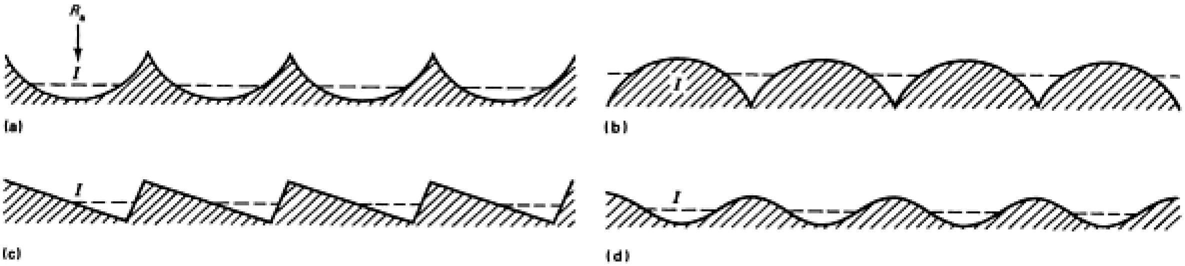


Fig. 3 Profiles for four different surfaces with the same roughness average, R_a, and wavelength but varying skewness, R_{sk}. (a) $R_{sk} > 0$. (b) $R_{sk} < 0$. (c) $R_{sk} = 0$. (d) $R_{sk} = 0$. The dotted line shows the position of the mean value.

Hybrid Parameters. Slope and curvature are two examples of quantities that combine the concepts of height deviation and lateral displacement and are thus termed hybrid parameters. They may be defined analytically or digitally in several ways. Hybrid parameters have been used in a number of areas of tribology, such as theories describing elastic contact (Ref 15) and thermal contact conductance (Ref 16).

Statistical Functions

More complete statistical descriptions of the properties of surface profiles can be obtained from statistical functions, such as those used in random process theory and time series analysis (Ref 17, 18, 19). Four important statistical functions are the amplitude density function or height distribution, the bearing area curve, the power spectral density, and the autocorrelation function. The definitions and applications of these functions are described in several sources (Ref 2, 6, 7, 17, 18, 19, 20, 21, 22, 23). The power spectral density and the autocorrelation function will be discussed in detail here.

The power spectral density (PSD) decomposes the surface profile into its spatial Fourier component wavelengths. It is given analytically by:

$$\text{PSD}(F) = \lim_{L \to \infty} \left(\frac{1}{L}\right) \left| \int_0^L y(x) \cdot e^{-2i\pi Fx} dx \right|^2 \quad \textbf{(Eq 4)}$$

and is estimated in digitized form by:

$$\mathrm{PSD}(F) = \mathrm{PSD}(k) = \frac{1}{N\Delta} \left| \sum_{j=1}^{N} y(j) \cdot e^{-2i\pi kj/N\Delta} \right|^2 \quad \text{(Eq 5)}$$

In Eq 4 and 5, Δ is the lateral point spacing (sampling interval) of the digitized data points, the total length of the profile, *L,* is equal to $N\Delta$, and the set of spatial frequencies, *F,* in the digitized PSD is given by *k/L,* where *k* is an integer that ranges from 1 to *N*/2. Calculation of the digital Fourier transform in Eq 5 can be greatly speeded by using fast Fourier transform (FFT) algorithms (Ref 24, 25).

Figure 4 illustrates the sensitivity of the PSD to the different characteristics of surfaces produced by different processes. Figure 4(a) shows the PSD plotted for a highly sinusoidal surface, a prototype of those available from the National Institute of Standards and Technology (NIST) as standard reference materials (Ref 26, 27). The Fourier amplitude at the fundamental frequency of 0.01 μm^{-1} (wavelength = 100 μm, or 10^6 Å) is the dominant feature in the curve, but imperfections in the sinusoidal nature of the surface are also evident from the presence of higher harmonics in the spectrum. By contrast, Fig. 4(b) shows a more random surface produced by the grinding process and its PSD (Fig. 4c) that has a fairly decaying, albeit somewhat randomized, distribution, with little evidence of periodic components.

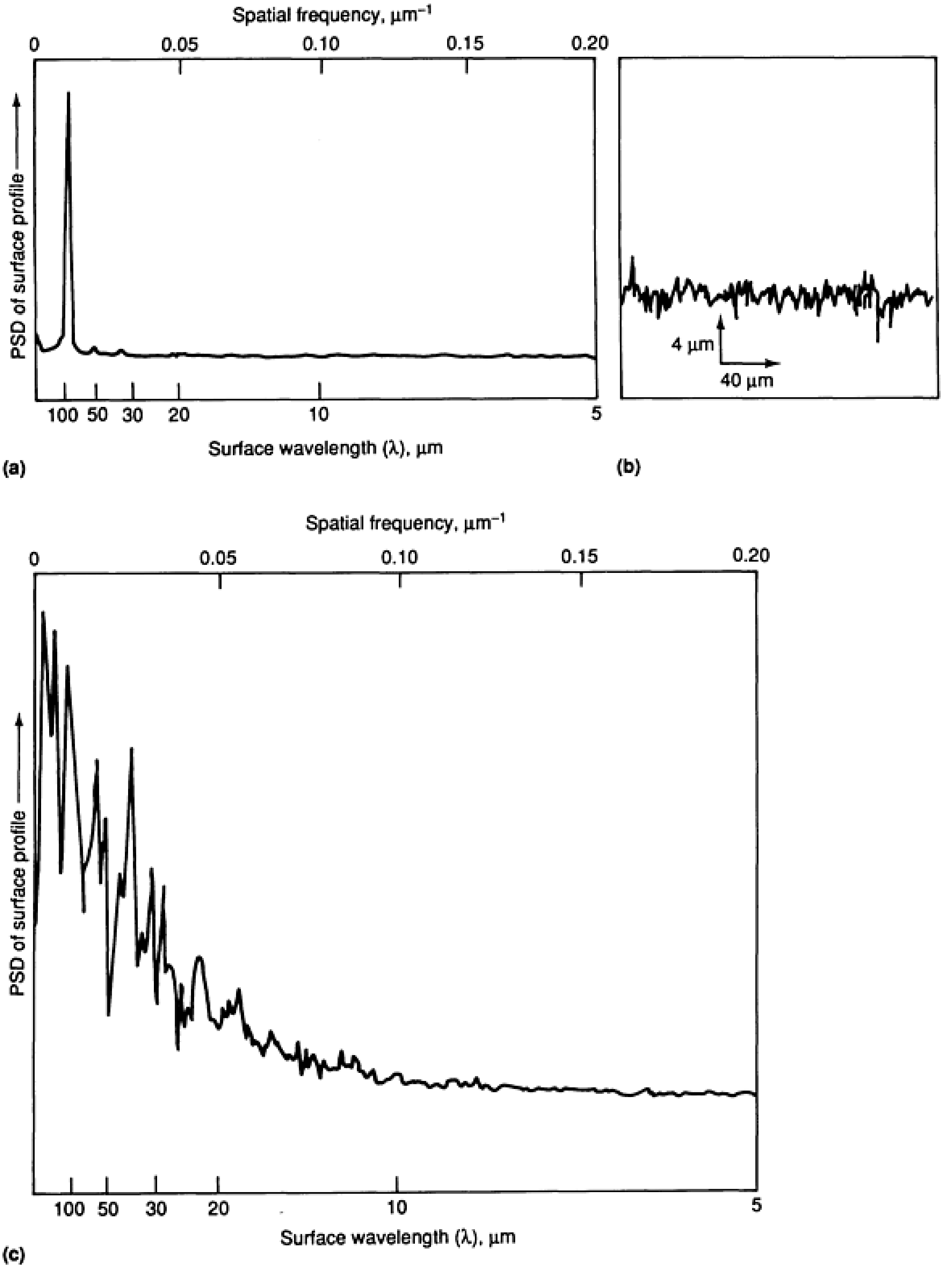


Fig. 4 Power spectral density functions for profiles of two roughness specimens. (a) Sinusoidal roughness specimen machined by diamond turning with R_a of 3 µm (120 µin.) and spatial wavelength of 100 µm (10^6 Å). (b) Roughness profile of a commercially ground surface whose PSD is shown in (c). (c) Commercially available specimen of a ground surface with R_a of 0.6 µm (24 µin.). Both power spectra are calculated for profiles having trace lengths of 3.8 mm (0.15 in.). The ordinate values are equal to the square root of the power spectral density values and are plotted in arbitrary units.

Autocovariance and Autocorrelation. The complementary function to the power spectral density is its Fourier transform, the autocovariance function, $C(\tau)$:

$$C(\tau) = \int_{-\infty}^{\infty} [\mathrm{PSD}(F)] e^{2i\pi F\tau}\, dF \qquad \text{(Eq 6)}$$

Alternatively, the autocovariance function can be calculated directly from the profile itself. That formula is given by an overlap integral of shifted and unshifted profiles:

$$C(\tau) = \frac{1}{L}\int_{0}^{L} [y(x)][y(x+\tau)]dx \qquad \text{(Eq 7)}$$

where the quantities L and $y(x)$ have been defined previously.

The value of the autocovariance function at zero shift ($\tau = 0$) is by definition equal to the mean square roughness, R_q^2, of the profile, provided an appropriate mean line has been subtracted from the profile to calculate $y(x)$. When the autocovariance function is normalized by dividing by the zero shift value, the result is known as the autocorrelation function (ACF), $c(\tau)$:

$$c(\tau) = \frac{C(\tau)}{R_q^2} = \frac{\int_{0}^{L} [y(x)][y(x+\tau)]dx}{\int_{0}^{L} [y^2(x)]dx} \qquad \text{(Eq 8)}$$

If the fact that the overlap between the shifted and unshifted profiles decreases as the shift distance increases for a finite length profile is taken into account and the digital formulation is simultaneously used, the ACF can be estimated by:

$$c(\tau) = \left(\frac{1}{N-j}\right)\left(\frac{\sum_{i=1}^{N-j} [y(i)][y(i+j)]}{\sum_{i=1}^{N} y^2(i)}\right) \qquad \text{(Eq 9)}$$

The autocovariance and autocorrelation functions are useful for visualizing the relative degrees of periodicity and randomness in surface profiles. For example, Fig. 5 shows ACFs for germanium and silicon surfaces calculated from surface profiles measured by a stylus instrument (Ref 28). Both surfaces were machined by diamond turning under similar conditions, but the periodicity imposed by the feed of the tool was much stronger on the silicon surface and thus its ACF is highly periodic. Because of its more random surface topography, the germanium surface exhibits a strongly decaying ACF, with only a small amount of periodicity shown as a barely visible oscillation.

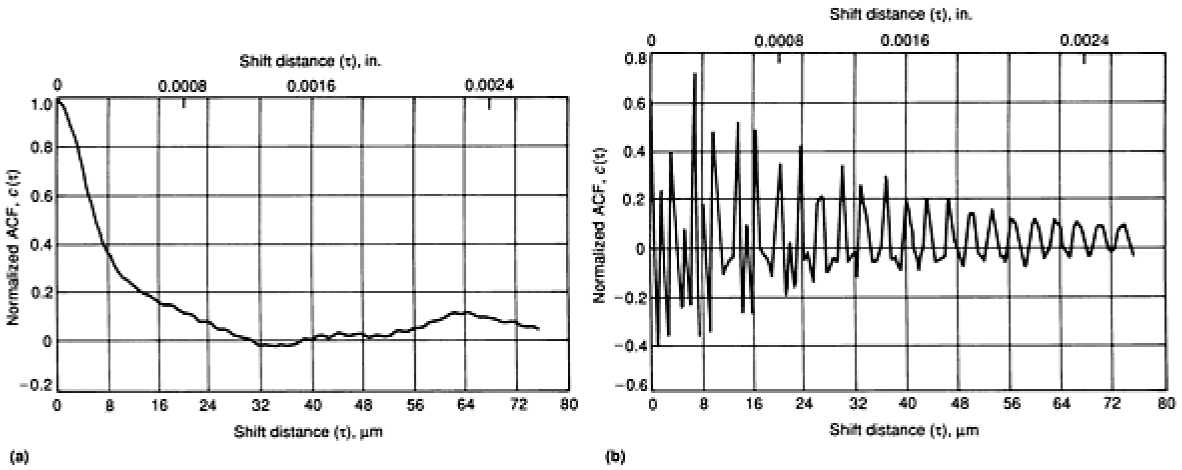


Fig. 5 Autocorrelation functions for single-crystal surfaces machined by diamond turning. (a) Germanium. (b) Silicon. The curve for germanium is an average of three ACFs measured at different positions on the surface. For silicon, the curve is an average of four ACFs.

Therefore, as measures of the lateral structure of surfaces, the PSD and the ACF seem to be useful in different ways. The PSD is useful for studying the strengths of various periodic components in the surface profile and for comparing these with the strength of the broad spectrum of random components. The ACF is useful for observing directly the lateral extent of the random structures on the surface by studying the decay in the function near zero shift.

Other Descriptors

As stated previously, scores of parameters and functions have been developed to quantify stylus profiles of surfaces, many of these for applications in tribology and engineering. The choice of roughness parameters and/or statistical functions to characterize the surface function adequately is critical for each application. An example from the automotive industry will be discussed here.

Finishing of engine cylinder bores by the plateau honing process (Ref 29) may result in an optimal cylinder bore surface texture. This surface has a negative skewness, R_{sk}, which results in good performance for running-in, long-term running, and lubrication of the cylinder bores. In such situations, the traditional roughness parameters, R_a, does not adequately estimate the performance of cylinder bore surfaces. The deep valleys, which can be used as oil reservoirs during engine operation, are better quantified by combining R_a with R_{sk}. Indeed, other parameters, such as the R_k family (Ref 30), have also been developed as functional descriptors of these types of surfaces.

Therefore, it may be possible to optimize the functional performance of engineering surfaces by combining the specification of surface texture parameters, material, and manufacturing process with the development of quality control procedures. This combination has been called "surface texture design" (Ref 31).

It is likely that in surface science, too, other topographic parameters can be developed that correlate well with performance. For example, Blakely and Somorjai (Ref 32) and others correlated the ability of a surface to foster certain surface chemical reactions with the presence of lattice steps on the surface. These ideas could be quantified in terms of a step density parameter that might be defined as the fraction of surface atoms that are adjacent to lattice steps.

Experimental Issues for Stylus Instruments

The height resolution and range of stylus instruments depend largely on two factors: the transducer and the mechanical noise and straightness of the traverse mechanism. In many research instruments, the transducer is a linear variable differential transformer (LVDT) (Ref 3) that, together with demodulation and amplification circuitry (Ref 33), produces an output directly proportional to surface height. Another type of transducer based on optical interferometry also produces a signal directly proportional to height (Ref 34). Inductive and piezoelectric transducers that are sensitive to the vertical motion of the stylus (that is, the time derivatives of the stylus tip height rather than the height itself) are also

available (Ref 33, 35). These types are generally more economical than LVDTs and are more generally used in industrial shops than in research or metrology laboratories.

An example of an LVDT is shown schematically in Fig. 6. As the stylus moves vertically, the ferrite core is displaced from the balance position of an alternating current (ac) bridge, to which the coils are connected, and which is driven at a high carrier frequency. The out-of-balance voltage signal that is demodulated and amplified is proportional to the vertical displacement of the stylus. Some transducers are so sensitive as to have rms noise levels of about 0.1 nm (Ref 7), which means that surface structures approximately one atom high can be detected. Other LVDT transducers have vertical range as large as several hundred micrometers or more, and, in general, the ratio of range to resolution is on the order of 10^5. One version of the optical transducer (Ref 34) mentioned above has an even wider range (~4 mm, or 0.16 in.) and a resolution of approximately 0.01 μm.

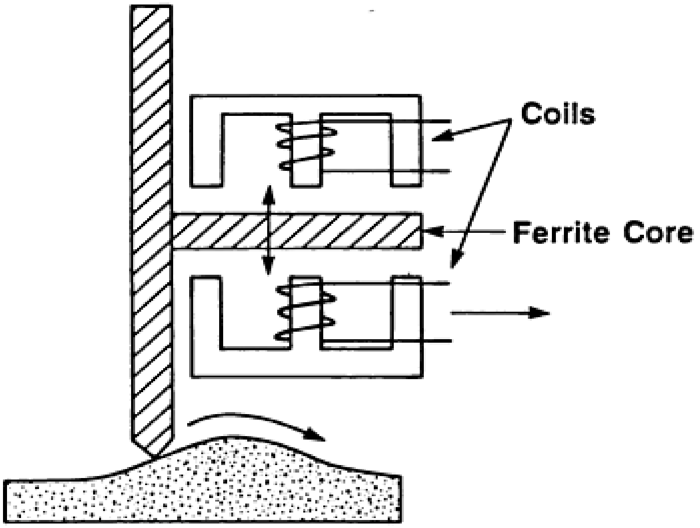


Fig. 6 Schematic showing key components of a high-sensitivity LVDT. As the stylus traverses the peaks and valleys of the surface, the ferrite core moves vertically with respect to the coils.

Three types of designs for achieving low scanning noise and good straightness of travel are shown in Fig. 7. The least expensive method involves the use of a skid (Fig. 7a), which traces a references line on the surface very near to the LVDT. The vertical position of the stylus is measured with respect to the path of the skid. The skid also serves as a high-pass mechanical filter because it averages over a fairly wide swath of surface peaks. Mechanical noise in the drive mechanism can be significantly reduced by this approach, because such noise does not couple well into relative motion between the stylus and the skid to which it is referenced.

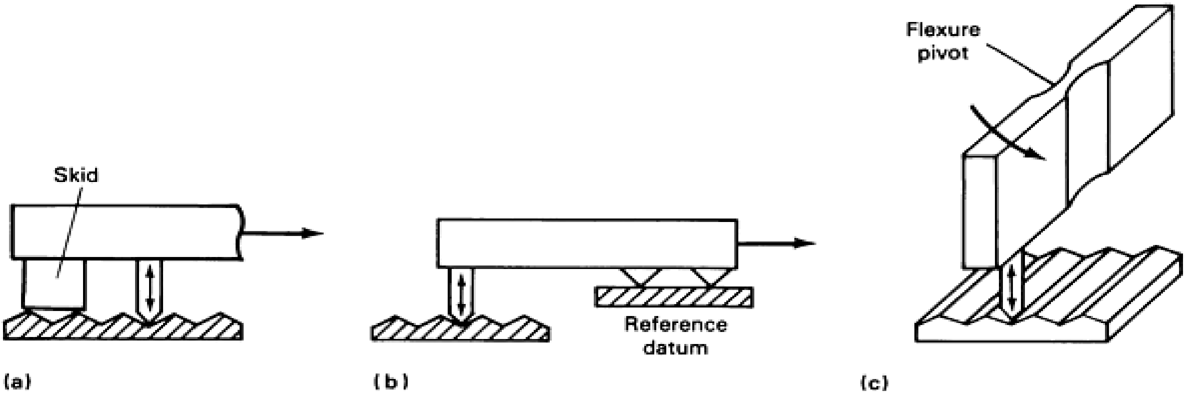


Fig. 7 Schematic showing three types of configurations used to constrain unwanted motion of the stylus. (a)

High-pass mechanical filter skid. (b) External reference datum to guide tracer motion. (c) Flexure pivot setup

Figure 7(b) schematically shows an external reference datum surface that guides the motion of the tracer. Well-engineered tracing systems may minimize the Abbé offset (Ref 36) between the reference surface and the measured surface by placing the stylus contact and the set of reference contacts as close to the same vertical line as possible. Alternatively, such systems may minimize the amount of angular motion and vibration in the carriage.

The design shown in Fig. 7(c) involves no sliding components at all. Rather, a flexure pivot or a set of pivots constrains the motion to a plane that is perpendicular to the axis of the pivot. The solid construction provides highly noise-free motion. One commercial design has been shown to provide approximately 0.1 nm rms noise (Ref 7) over a trace length of 1 mm (0.04 in.). Another variation of the flexure approach (Ref 37) involves multiple sets of flexures to allow for motion in two directions.

For measurements of smooth surfaces that require high vertical magnification, the straightness error of the traverse mechanism can seriously distort the measured profile. Figure 8(a) shows a distorted profile of a 91 nm (3.6 µin.) step-height specimen (Ref 31). The use of the skid can substantially eliminate this error (see Fig. 8b), because the reference datum is a portion of the surface located very close to the stylus. This consideration is closely related to the Abbé principle in dimensional measurement (Ref 36). The skid approach works well as long as the skid is traversing on a smooth reference surface. However, use of the skid can lead to unwanted distortions in the profile, particularly in cases where the surface contains isolated peaks (Fig. 9). Another problem with skids is the potential for surface damage, because the skid loading on the surface may be hundreds of times larger than the stylus force.

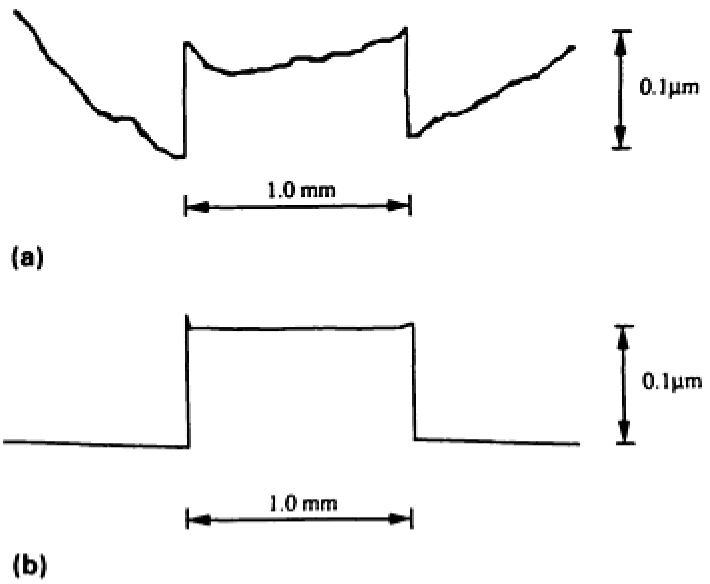


Fig. 8 Profile of a 91 nm (3.6 µin.) step-height specimen distorted by the straightness error of the traverse mechanism of the stylus instrument. (a) Stylus without skid. (b) Stylus with skid

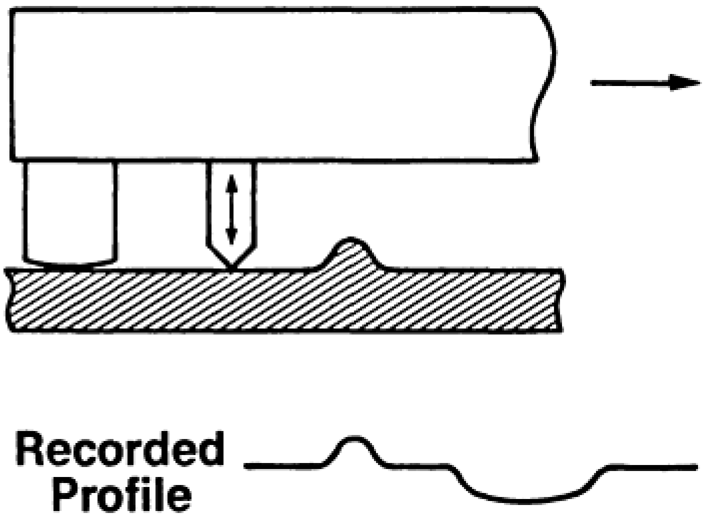


Fig. 9 Schematic showing an isolated peak on the specimen surface generating a false valley in the profile that was measured by a stylus instrument incorporating a skid

Lateral Resolution and Range. The draw-back of flexure designs is that the trace length is limited by the degree to which the flexure may be bent before the elastic bending forces become too large. In the two designs discussed above, that trace length limit is about 2 to 3 mm (0.08 to 0.12 in.). However, as a rule, lateral range is not a problem for stylus instruments. The lead-screw drive mechanisms of many instruments permit trace lengths on the order of 100 mm (4 in.) or more.

Because a stylus profile represents the convolution of the surface structure with the end form of the stylus, the lateral resolution depends critically on the size of the stylus tip and to a lesser degree on the flank angle. Styli are often flattened at the end, the width across the flat essentially determining the lateral resolution. A commonly available width is 2 μm (80 μin.), but pyramidal styli with tips as small as a nominal 0.1 × 0.1 μm (4 × 4 μin.) are also used. In some instances, highly spherical styli are also produced. Bennett and Dancy (Ref 7) and Vorburger *et al.* (Ref 38) show micrographs for these. Figure 10(a) shows a stylus with a tip width of ~0.05 μm as measured by the razor blade tracing technique (Ref 38, 39). One year later, after dozens of traces, the profile graph of the same stylus (Fig. 10b) shows a certain amount of wear, because the width has increased to approximately 0.15 μm (6 μin.).

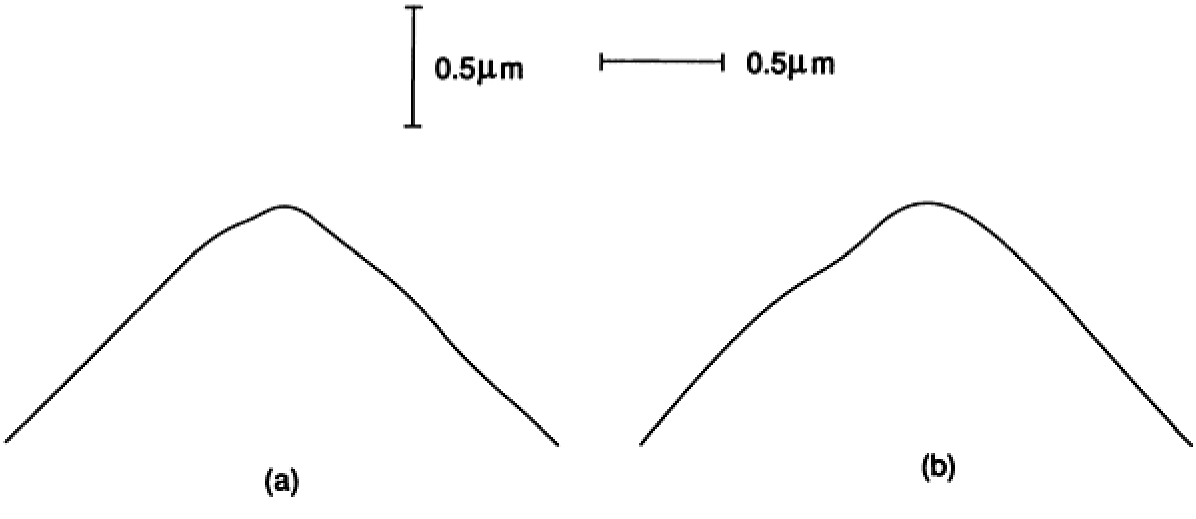


Fig. 10 Profile graph of low-usage stylus related to profile graph of identical stylus after 15 months of service. (a) Measured in March 1988 with a tip width of approximately 0.05 μm (2 μin.). (b) Measured in June 1989;

width is approximately 0.15 μm (6 μin.) after dozens of traces.

If the stylus tip is spherical, the lateral resolution can be smaller than the tip radius, *r*. Figure 11 shows the geometrical construction when a spherical stylus first contacts a step, which is assumed to be infinitely sharp. If $h \ll r$, the point of initial contact occurs at a position $\sqrt{2hr}$ to the left of the step edge; thus, $\sqrt{2hr}$ represents the width of the transition and the effective resolution of the stylus. The calculation assumes that no deformation occurs at the single point of contact. In the case of a 1 μm (39 μin.) radius stylus and a step of height 10 nm, the calculated transition width is 0.14 μm (5.5 μin.).

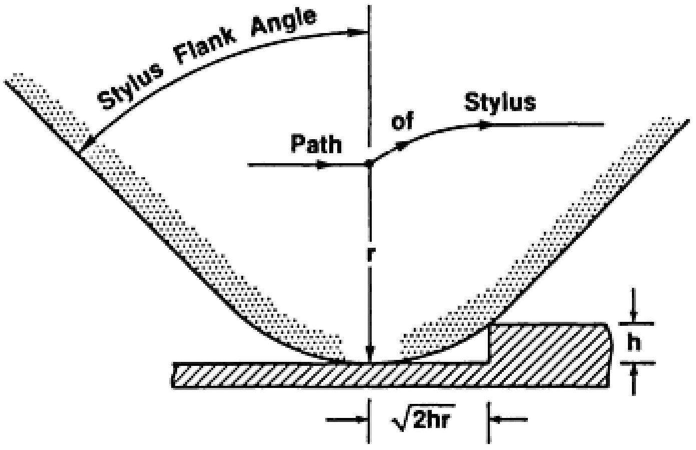


Fig. 11 Schematic showing the path traced by a spherical stylus tip traversing over a small, sharp step

The flank angle of the stylus (Fig. 11) also affects its resolution. The stylus cannot accurately profile over slopes greater than the flank angle (usually ±45°). This is not a problem for most types of smooth surfaces where the average surface slopes are only a few degrees, but it does complicate measurements of surfaces with steep cracks or holes, such as those on ceramics or cast materials.

An example of the lateral resolution capability of a stylus instrument is illustrated in Fig. 12(a) (Ref 39), which shows a profile of a 6000 lines/mm grating measured using a flexure-pivot-type stylus instrument with a 0.05 μm (2 μin.) tip width (Fig. 10a) and 0.6 μN stylus loading. A piezoelectric stage was used to drive the surface under the stylus to obtain the high lateral magnification. By contrast, Fig. 12(b) shows a profile of the same surface with a much wider (0.5 μm, or 20 μin.) stylus tip.

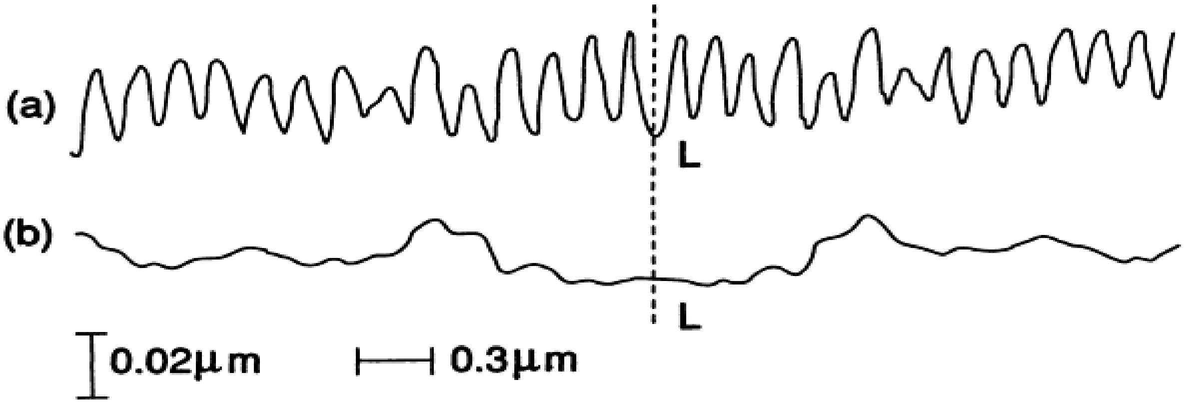


Fig. 12 Surface structure of the 6000 lines/mm grating recorded by a flexure-pivot-type stylus under a 0.6 μN

load. (a) Measured by a 0.05 μm (2 μin.) width stylus. (b) Measured by a 0.5 μm (20 μin.) width stylus

To summarize, the convolution of the stylus with the surface reduces the widths of valleys and increases the widths of peaks in the output profile. The minimum detectable valley width is roughly equal to the stylus tip width itself, but may be smaller for spherical styli tracing over very shallow valleys. From geometrical arguments, the minimum detectable peak width can be quite small in principle, but if it becomes considerably smaller than the stylus tip width, the area of contact would be decreased, the stylus pressure would increase, and the stylus might tend to deform such an asperity. However, peak spacings of 0.05 μm (2 μin.) have been detected (Ref 39) on the surfaces of random-profile precision roughness specimens (Ref 40) using commercially available stylus tips.

Bandwidth Limits for Surface Metrology. The foregoing discussion leads to the important concept of bandwidth limits (Ref 28) for surface metrology. Stylus instruments are sensitive to a certain bandwidth of surface spatial frequencies, limited at the high-frequency end by the stylus width and the sampling interval and at the low-frequency end by the traverse length or by high-pass electrical filtering. One of the most important considerations in stylus profiling is whether or not the wavelength domain of interest on the measured surface falls in the flat portion of the transmission function of the instrument. For example, at position X in Fig. 13 (Ref 41), there is essentially no attenuation caused by the factors that limit either range or resolution. One type of instrument has a traverse of 2.5 mm (0.1 in.) and can employ stylus tips having widths as small as 0.1 μm (4 μin.). Therefore, the effective band-width of spatial wavelengths sensed by this instrument extends over the range of approximately 0.1 to 2500 μm (4 to 10^5 μin.) for its unfiltered operation. For many industrial applications,the standard stylus has a radius of 10 μm (400 μin.) and the long-wavelength cutoff (λ_c) is determined by a standard high-pass electrical filter that is often set to 800 μm (3 × 10^4 μin.). The spatial wavelength bandwidth of these instruments extends approximately over 10 to 800 μm (400 to 3 × 10^4 μin.). Spatial wavelengths larger or smaller than this range are detected with reduced or nil sensitivity.

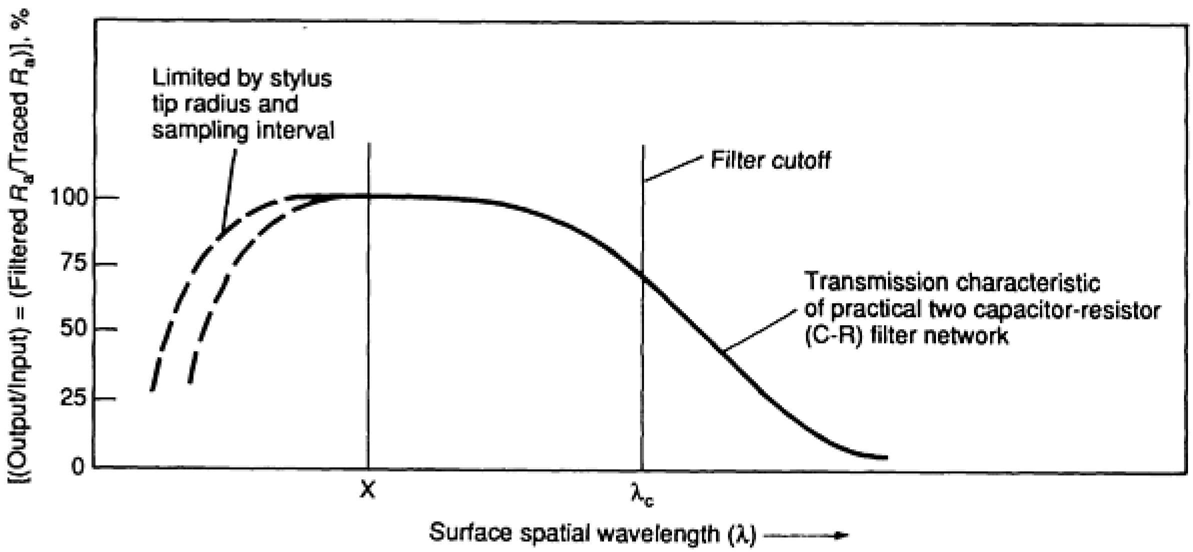


Fig. 13 Transmission characteristics of a stylus instrument as a function of surface spatial wavelength. Source: Ref 41

It is important for the operator to use a bandwidth that encompasses the spacing of the surface roughness structures to be evaluated. This means that, on the one hand, the finest surface features to be assessed must be more widely space than the stylus tip width or radius and, on the other hand, that the nominal instrument cutoff must be several times longer than the most widely spaced features to be assessed.

However, the bandwidth also affects the arms noise of the measurement. The wider the bandwidth, the more noise enters into the output profile. The noise contribution to a roughness measurement can be checked for most stylus instruments by measuring an optical flat, because such a specimen ordinarily has a roughness that is smaller than the noise resolution of the instrument. The resulting value of measured R_a or R_q represents the noise resolution of the instrument, including both the mechanical noise of traversing and the electrical noise over the finite spatial bandwidth as discussed above.

Stylus Load and Surface Deformation. The logical parameters that determine whether surface damage will be caused by stylus load are the surface hardness, the stylus force, the stylus tip width, and, to a lesser extent, the stylus speed. A stylus tip width of 1 μm (40 μin.) should not produce detectable damage on metal surfaces as soft as gold as long as the stylus force is smaller than about 0.03 mN.

Many types of stylus instruments use stylus forces of 0.5 mN and higher, but these are normally used with stylus tip sizes on the order of 10 μm (400 μin.). Because the pressure is inversely proportional to the area of contact, the pressure on the surface caused by stylus loading is smaller for a 10 μm (400 μin.) stylus with a 0.5 mN force than it is for a 1 μm (40 μ in.) stylus with a 0.03 mN force. Even if the stylus leaves a visible track, the resulting profile is likely to be accurate, because the variation in the depth of the track over the surface should be significantly smaller than the depth itself. However, if a skid is used for stylus profiling, the measured surface can be seriously damaged by the skid, whose loading is hundreds of times larger than the stylus loading.

The above discussion pertains only to plastic or irreversible deformation of the surface by stylus loading. Characterizing the elastic or reversible deformation (Ref 6) is much more difficult, but the elastic deformation is expected to be very small (Ref 42).

In a study of plastic damage, Song and Vorburger (Ref 39) measured a 2160 lines/mm gold grating with a 0.5 μm (20 μ in.) stylus tip width. When the stylus loading increased from 0.6 to 100 μN, the grating profile in the same position was attenuated (Fig. 14a-d). When the stylus loading was reduced again to 0.6 μN (Fig. 14e), most of the periodic structure of the profile in Fig. 14(a) had been plastically eliminated by the previous loading conditions and did not reappear. However, a few of the fine peaks did reappear, and the difference between the profiles in Fig. 14(d) and 14(e) suggests that some features were only elastically deformed by the increased stylus loading.

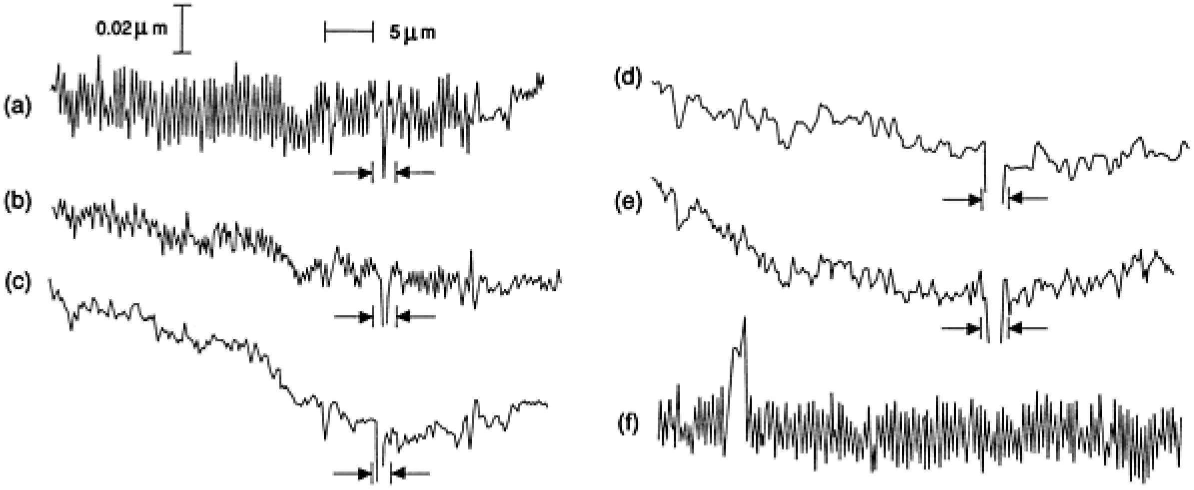


Fig. 14 Effect of stylus loading on the surface of a gold grating with 2160 lines/mm. Nominal stylus radius of 0.5 μm (20 μin.). Stylus loading: (a) 0.6 μN. (b) 25 μN. (c) 50 μN. (d) 100 μm. (e) 0.6 μN. (f) 0.6 μN, different position

Other Distortions. Stylus flight (Ref 43, 44, 45) and profile digitization are two other sources of profile distortion. Stylus flight can occur when the stylus encounters a sharp change in the surface topography, such as a steeply rising surface step. The logical parameters that affect this phenomenon are the stylus speed, the stylus force on the surface, the stylus tip size, the damping constant in the vertical direction, and the rate of change of the surface slope. A key tradeoff occurs between stylus force and speed. A magnetic phonograph cartridge with a force of 20 mN can have a record disk traverse beneath it at a tangential speed of 500 mm/s (20 in./s) without losing contact, but a stylus with a force of 0.5 mN must travel much more slowly, about 1 mm/s (0.04 in./s), to maintain contact. The usual symptom of stylus flight is a peak in the measured profile with a sharp rise and slower tail occurring after the stylus encounters a sharp peak on the surface. The accuracy of such features can be verified by remeasuring the same profile at a slower speed.

Stylus profiles are routinely digitized for the purposes of computer processing and mass storage. In order to obtain an accurate digital representation of the profile, the peak-to-valley height of the profile should consist of many vertical

quantization levels, and the widths of the surface features to be studied should consist of many lateral sampling intervals. In addition, if a distribution of surface peaks and valleys is being characterized, there should be enough points in the profile to give an adequate statistical sampling of the variability of these structures. A system at NIST used 4096 vertical quantization levels and 4000 digitized points. These values seem to provide adequate resolution for many applications.

Finally, a ubiquitous source of confusion is simply the difference between the horizontal and vertical magnification of surface profile records. The ratio of vertical to horizontal magnification can be 100:1 or higher in some applications. This effect is not a source of error, but leads to misperceptions of the true appearance of surface texture because the resulting profile records have highly sloped and sharply peaked structures. Figure 15, taken from Reason (Ref 33), shows a comparison between a profile measured with a 1:1 ratio and one with a 25:1 ratio. The qualitative impressions derived from the two pictures are quite different. In reality, surfaces are much less jagged than they appear from conventional profile records.

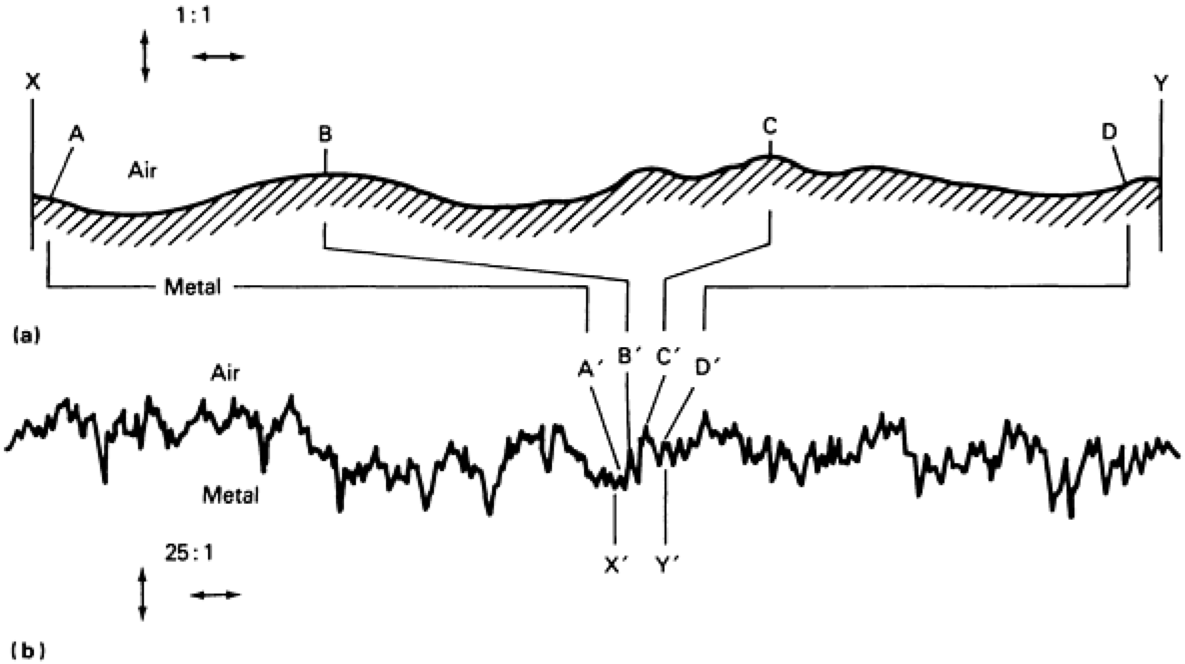


Fig. 15 Stylus profiles obtained with two different aspect ratios. (a) Undistorted 1:1 representation. (b) Plot in which the horizontal scale has been compressed by a factor of 25 with respect to the vertical scale. Source: Ref 33

Examples of Roughness Measurement Results. Of all the surface profiling concepts discussed in this article thus far, the most widely used output parameter is a roughness average, R_a, and perhaps the most important instrument parameter is the long-wavelength cutoff. A few measurement results for R_a from typical metal finishing processes will now be discussed, with the instrument cutoffs noted as well.

The surfaces of metal components can be finished by any of a number of different processes. Typical ranges for the roughness average achieved by a large number of processes are given in Ref 4. The ranges of measurements made by the authors for a few of these processes are shown in Table 1, along with the long-wavelength cutoffs used. These results represent the highest and lowest values that were obtained on roughness comparison surface replicas for each type of finishing process. Nearly all the replicas are commercially available.

Table 1 Extremes of arithmetic average surface roughness, R_a, as a function of selected metalworking finishing processes

Finishing process	Measured values of surface roughness[a]							
	Minimum				**Maximum**			
	R_a		**Cutoff**		R_a		**Cutoff**	
	μm	**μin.**	**mm**	**in.**	**μm**	**μin.**	**mm**	**in.**
Ground	0.024	0.96	0.8	0.03	3.0	120	0.8	0.03
End milled	1.4	56	0.8	0.03	11	440	No cutoff	
Side milled	1.2	48	2.5	0.10	14	560	No cutoff	
Shaped or turned	0.6	24	0.8	0.03	18	720	2.5	0.10
Electrical discharge machined (EDM)	0.4	16	0.8	0.03	7.5	300	0.8	0.03
Cast	0.9	36	0.8	0.03	72	2900	16	640

(a) For various finishing processes as measured and recorded by J.F. Song and T.V. Vorburger between 1976 and 1991. These values do not necessarily represent the entire range of values obtainable by these processes.

The cutoffs were chosen either to be several times longer than the typical spacing produced by the surface finishing process or to be 0.8 mm (0.03 in.) as a minimum. In general, the spacing of the machining marks increases with roughness; therefore, for the same finishing process, rougher surfaces require longer instrument cutoffs.

Ceramic materials are being increasingly used in industrial machinery. Although surface finishing processes are more expensive for ceramics than for metals, the ranges of roughness values achievable for both materials are generally similar. However, many types of ceramic surfaces are porous, and thus the finished surface is characterized by fairly smooth plateaus and deep holes. Therefore, values of skewness tend to be negative, and the values of peak-to-valley parameters tend to be larger relative to R_a for ceramic surfaces than for metal surfaces.

Instrument Calibration

Tribologists often make comparisons of surface texture to determine the existence, extent, and causes of surface wear. These comparisons can be confused by differences in surface measurements taken under different conditions. Are these differences caused by the measuring instruments, the measured surface, or the variation of measuring conditions? How can surface measurements be made accurate and when can they be compared? These questions involve both instrument calibration, correct measuring procedures, and the use of various calibration and check specimens.

General Calibration Issues. The measurement conditions that should be defined, calibrated, or checked for a stylus instrument are (Ref 31, 41, 46):

- Magnification, both in the vertical and horizontal directions
- Stylus tip
- Stylus loading
- Type of skid or reference datum
- Type of filter, reference line, and cutoff length
- Profile digitization
- Algorithms for calculating parameters
- Number and distribution of profiles on the surface

Four types of calibration specimens can be used for this purpose according to ISO standard 5436 (Ref 41): step-height specimens for calibrating the vertical magnification, specimens with fine grooves for checking stylus condition, specimens with periodic profiles for checking vertical and horizontal magnification as well as the character of an electronic filter, and specimens with random profiles for checking the overall response of an instrument (Ref 31, 41).

The vertical magnification of a typical commercial stylus instrument is generally accurate to 10% or better, depending on the fineness of the application. For accurate dimensional measurement of surface structures, the instrument must be calibrated. This is often done by measuring the recorded displacement produced by traversing a step whose height has

been calibrated by interferometric measurement. Calibration in the vertical direction becomes difficult at very high magnifications where the desired resolution may be at the nanometer or subnanometer level, somewhat beyond the resolution capabilities of conventional interferometric techniques. In that case, interferometric techniques that incorporate electronic phase measurement (Ref 47) constitute one approach to providing calibrated measurements of small step heights. The sources of uncertainty in surface height calibration and estimates of their magnitudes are discussed elsewhere (Ref 31, 46, Ref 48).

In the lateral direction, the relative displacement of the stylus over the surface can be measured directly by a laser interferometer (Ref 37). Alternatively, calibrated grids or other types of periodic surface specimens (Ref 26) can be used as secondary displacement standards.

Comparison of Roughness Parameters. In order to make surface measurements results comparable, the measurement conditions mentioned above should be precisely defined and specified, especially the stylus size and cutoff length, which limit the bandwidth of the measured profile. The accuracy of surface measurements of manufactured parts is aided further by a well-established measurement procedure, such as the following (Ref 31):

1. Calibrate the vertical magnification of the instrument using a step specimen whose calibrated step height covers the range of surface heights of the engineering surfaces to be measured
2. Verify that the calibration was correct by measuring either the calibrated step height again or a roughness specimen with calibrated R_a, such as a sinusoidal specimen (Ref 27)
3. Measure the engineering surfaces of interest
4. Check the measurement by measuring a check specimen with a waveform identical or similar to that of the measured surface. The R_a or other roughness parameter value of the check specimen should have been calibrated under the same measuring conditions with the same instrument characteristics as the measurement in step 3

In addition, the instrumental parameters, such as filter setting, stylus loading, and straightness of the mechanical motion, should be checked periodically.

Existing roughness calibration specimens can be used as check specimens for a wide range of engineering surface measurements. For example, when the measured engineering surfaces have highly periodic profiles, such as those obtained by turning, planning, or side-milling processes, periodic roughness specimens with triangular, cusped-peak, or sinusoidal profiles can be used as check standards. When the measured engineering surfaces have random profiles, as obtained by grinding, lapping, polishing, and honing processes, the random roughness specimens originating from the Physikalisch Technische Bundesanstalt in Germany (Ref 49) or the Chang Cheng Institute of Metrology and Measurement in China (Ref 40, 50) can be used. These sets combined would cover the range of R_a values from 1.5 to 0.012 μm (59 to 0.5 μin.). If the checking measurement shows that the difference between the measured result for the check specimen and its certified value under reference conditions was within a given tolerance, the measurement of the engineering surface is considered to be under good quality control (Ref 31).

In tribology experiments, if surfaces measured under identical conditions are being compared, the instrument is needed only as a comparator and its absolute calibration is of secondary importance. In such case, only a pilot specimen may be needed for surface measurement quality control. The pilot specimen could be selected from the measured engineering parts or could be an engineering surface with the same surface texture pattern and a similar roughness parameter value as the test surfaces, produced by the same manufacturing process. It should also have good surface texture uniformity. The stylus instrument should be checked for measurement repeatability by measuring the same trace approximately 15 to 20 times. After that, several measurements should be made daily at positions evenly distributed in a small measuring area designated on the surface of the pilot specimen. The user should then be able to detect a significant change in the characteristics of the instrument.

Comparison of the surface profiles often yields more useful information in tribology experiments that the simple comparison of roughness parameters. However, profile comparison requires that the tested surface be relocated in the exact same place from one measurement run to the next. Discrete, recognizable surface features, either natural or artificial, could be used for relocation. In Fig. 14, for example, a deep valley on the measured gold grating surface (see arrows) provided a means of orienting these profile graphs from run to run.

Applications

Metalworking. Measurements of surface roughness for metalworking components likely form the bulk of surface roughness measurements throughout the world. The automotive industry is one example where the manufactured surfaces are carefully specified. Table 2, now about 16 years old (Ref 13), shows roughness specifications in terms of the roughness average, R_a, for a number of automobile components. It is likely that these specifications were drawn up empirically and were probably similar to specifications elsewhere in the automotive industry. However, there is no real collective body of knowledge that describes these types of specifications and the reasons for them. As far as can be told, the information is scattered throughout the literature or is proprietary.

Table 2 Typical surface roughness specifications of 1976 model year automotive engine components

Components	Manufacturing process	Car No. 1		Car No. 2	
		μm	μin.	μm	μin.
Cylinder block					
Cylinder bore	Honing	0.41-0.51	16-20	0.51-0.64	20-25
Tappet bore	Reaming	1.5-1.9	60-75	2.0-3.0	80-120
Main bearing bore	Boring	1.5-2.0	60-80	3.3-3.8	130-150
Head surface	Milling	1.0-1.3	40-50	4.8-5.3	190-210
Piston					
Skirt	Grinding-polishing	1.1-1.4	45-55	1.0-1.3[(a)]	40-50[(a)]
Pin bore	Grinding/polishing	0.76-0.97	30-38	0.28-0.33[(a)]	11-13[(a)]
Piston pin	Grinding-lapping	0.23-0.30	9-12	0.08-0.13	3-5
Crankshaft					
Main bearing journal	Grinding-polishing	0.10-0.15	4-6	0.15-0.23	6-9
Connecting rod journal	Grinding-polishing	0.10-0.15	4-6	0.15-0.23	6-9
Camshaft					
Journal	Grinding-polishing	0.10-0.15	4-6	0.36-0.46	14-18
Cam	Grinding-polishing	0.38-0.51	15-20	0.56-0.66[(a)]	22-26[(a)]
Rocker arm					
Shaft	Grinding	0.36-0.51	14-20	0.51-0.56	20-22
Bore	Honing-polishing	0.74-0.81	29-32	0.76-1.0	30-40
Valves					
Stem:					
Intake	Grinding	0.86-0.97	34-38	0.41-0.56	16-22
Exhaust	Grinding	0.46-0.51	18-20	0.36-0.51	14-20
Seat:					
Intake	Grinding	0.64-1.0	25-40	0.76-1.0	30-40
Exhaust	Grinding	0.86-1.1	34-45	0.76-0.89	30-35
Tappet					
Face	Grinding	0.10-0.13	4-5	. . .	
Outside diameter	Grinding	0.36-0.46	14-18	. . .	
Hydraulic lifter					
Face	Grinding-polishing	0.56-0.64	22-25	0.38-0.51[(a)]	15-20[(a)]
Outside diameter	Grinding-polishing	0.36-0.41	14-16	0.33-0.36[(a)]	13-14[(a)]

Source: Ref 13

(a) Grinding only; no polishing.

Griffiths (Ref 51) attempted to systematize some of the knowledge on surface function. Table 3, taken from his paper, lists the correlations between surface physical properties and various causes of component failure. The circles are taken from previous work of Tonshoff and Brinksmeier (Ref 52) and the squares from Griffiths' additional research. The surface texture influences failure occurring by plastic deformation, fatigue, and corrosion. Griffiths also listed the influence of surface parameters on component performance (Table 4). This table discusses not only roughness and waviness, but also the metallurgy and chemistry of the surfaces and other qualities as well. Roughness is particularly important for sealing, dimensional accuracy, preserving the cleanliness of the component, optical reflectivity, and several other functions.

Table 3 Effect of surface properties on component failure causes

Cause of failure	Surface physical properties[a]						
	Yield stress	**Hardness**	**Strength**	**Fatigue strength**	**Residual stress**	**Texture**	**Microcracks**
Plastic deformation	•	•				□	
Scuffing/adhesion		•					
Fracture/crack			•				[ocir]
Fatigue				•	[ocir]	[ocir]	•
Cavitation		[ocir]					[ocir]
Wear		•		□	[ocir]		□
Diffusion						○	
Corrosion		□			[ocir]	□	•

Source: Ref 51

(a) From original 1980 survey: •, strong influence; [ocir], traceable influence; ○, supposed influence. Later survey: □, Traceable influence.

Table 4 Effect of surface parameters on component performance

Performance parameter	Surface parameter[a]							
	Roughness	**Waviness**	**Form**	**Lay**	**Laps and tears**	**Chemistry**	**Metallurgy**	**Stress and hardness**
Sealing	•	○		○	○			•
Accuracy	•	○	○		○			•
Cleanliness	•				○	•		
Reflectivity	•	○		○	○	○		
Tool life	•			○	•		•	•
Load carrying	○	•						○
Creep	○					•	○	•
Magnetism						•	•	•
Electrical resistance	•					○	○	
Assembly	○				○	•		○
Fluid flow	•	○		○				
Joints	•	○		○				•

(a) •, strong influence; ○, supposed influence.

Tribology and Wear. An important research direction in tribology is to determine the relationship between surface texture and wear properties, and the variation of surface texture during the water process. Many investigators use standard test geometries for wear and friction tests, such as pin-on-disk or four-ball tests (Ref 53, 54, 55, 56). The amount and structure of damage to these compounds is of great interest in such tests. Key measurable parameters are the volume of material removed by wear and the surface area of the water scar.

As discussed by Whitenton and Blau (Ref 55), both two-dimensional analysis of profiles and three-dimensional analysis of surface topography maps can be used to assess wear damage. In the two-dimensional approach, a profile of the wear scar is obtained and the area lost or gained in the wear region is estimated. By projection, the volume can be estimated as well. The profile can be obtained by stylus measurements or by image analysis of the scar.

In the three-dimensional approach, the measurement system generates a matrix of X, Y, and Z values that describe the topography of the surface after the test. Parameters such as surface area can be determined from this matrix. In addition, the volume removed by wear can be obtained by comparing the surface map with that for the unworn surface. An

important advantage of this method is its accuracy; it produces the most direct measurement of the wear volume. One disadvantage is that it is more time consuming than the two-dimensional method.

Figure 16 shows the surface topography that resulted from measuring a bottom ball in a four-ball test (Ref 56) that used 6.35 mm (0.25 in.) radius α-alumina balls. These three-dimensional data of the wear scar surface were carefully filtered to remove extraneous instrumental errors.

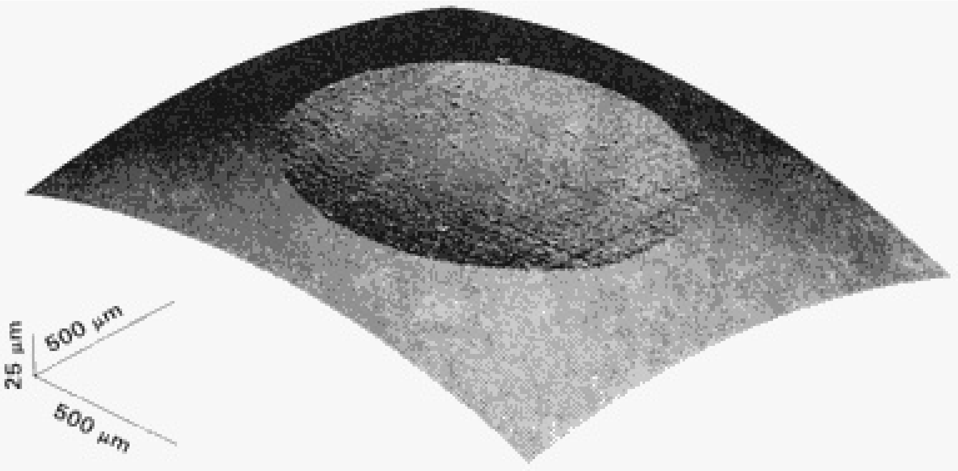


Fig. 16 Bottom-ball topographic data for a four-ball test showing a round wear scar. Source: Ref 56

Figure 17 shows the relationship between the wear volume of the top ball scars and the bottom ball scars in the four-ball test. Five sets of balls were tested at room temperature while immersed in paraffin oil. Because there are three bottom balls which were simultaneously, three times the wear volume for one ball is plotted along the *x*-axis. The scar volume of the top ball is plotted along the *y*-axis. Under the five different sets of experimental conditions, the total wear volume lost for the bottom ball scars as calculated from surface profiling appears to be about equal to the wear volume lost for the top ball scar.

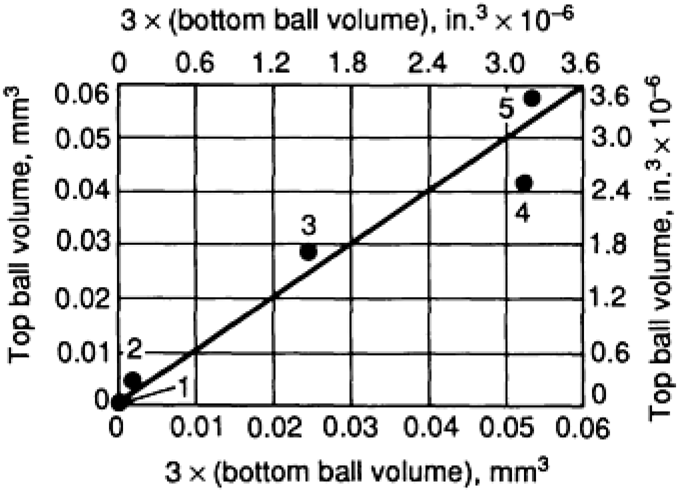


Fig. 17 Relation of the wear volumes of the top-ball wear scars to the bottom-ball wear scars. Because there are three bottom balls, three times the wear volume for one ball is plotted along the *x*-axis. The top-ball scar volume is plotted along the *y*-axis. A 1:1 45° line is also drawn. The numbered data points correspond to the test numbers. Source: Ref 56

Another application of surface texture measurements in tribology is the examination of used components to gain information on the wear mechanism (Ref 57). For example, the mechanism of scuffing involves the destruction of surfaces by the welding and fracture of asperity contacts. Such surfaces are easily distinguished from those produced by controlled running-in wear. Many engines use specially formulated "first-fill" lubricants designed to assist the running-in of the surfaces. This running-in is crucial for obtaining satisfactory service life.

Bovington (Ref 57) has observed how a properly run-in surface can be distinguished from a scuffed surface. Generally, the run-in surface contains a number of flat plateaus, the peak-valley roughness is about half that of the new surface, and the skewness, R_{sk}, is negative. Running-in proceeds in a controlled manner, that is, with the truncation of surface peaks but without abrasive or adhesive wear processes. The truncations or plateaus result in a reduction of the contacts pressures, and their presence is a good indication of long service life. Scuffing, on the other hand, generates new surfaces. Therefore, the peak-valley roughness does not decrease, and the R_{sk} parameter does not become progressively more negative.

Bovington (Ref 57) has also observed that modern engine design and lubrication technology are so advanced that the old methods of evaluation of wear, such as weight loss, are becoming irrelevant. The lubricant industry needs to begin defining wear in terms of changes in surface texture.

Davis *et al.* (Ref 58) measured the three-dimensional topography of various places in a honed engine cylinder bore and related the topography to component wear. Based on detailed results, they developed a chart showing oil volume in cubic millimeters versus the amount of the surface that would be truncated by the wearing process. The oil volume is related to the volume of the surface valleys, calculated from their three-dimensional topographic measurements. Their mathematical truncation process was a simulation of a wear process that cuts off the surface peaks. For an engine cylinder bore, oil volume is of crucial importance. Figure 18 shows data taken from the analyses of one of their three-dimensional topographic maps. As the truncation proceeds, the oil volume in the valleys decreases. Based on their information and measurement results, Davis *et al.* (Ref 58) predicted that the component would begin to fail at a truncation level between 60 to 70%, because the oil volume would decrease to unacceptable levels.

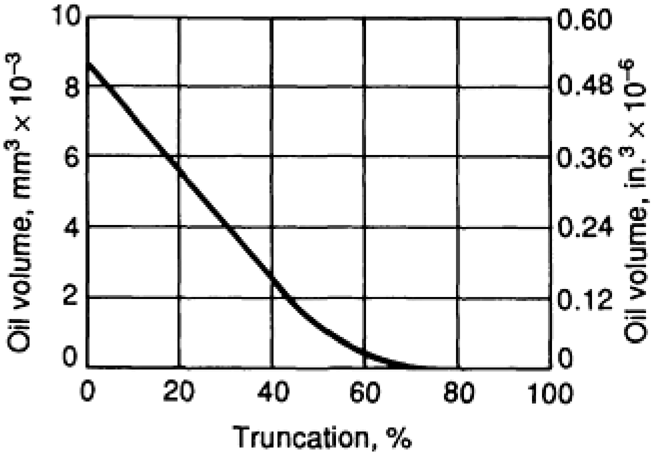


Fig. 18 Oil volume for cylinder bores estimated by mathematical truncation of a surface topography map. Source: Ref 58

Magnetic Storage. Tribology is especially important to the functioning of tapes and disks in the magnetic recording industry (Ref 59), including the hydrodynamic properties of flying read heads, the lubrication of tapes and disk, and the sliding contact between a disk and a read head upon startup. Surface roughness is also important. Figure 19 shows results from Bhushan *et al.* (Ref 59, 60) for the measured coefficient of friction of six CrO_2 magnetic tapes sliding against a glass head as a function of the rms roughness measured with an optical profiler. The tapes all had the same composition; the variation in rms roughness was achieved by using different calendering pressures during the finishing process. The coefficient of friction decreased rapidly up to an rms roughness of about 40 nm,then seemed to remain fairly level. However, when the friction results were plotted versus the real area of contact (normalized to the applied load), an

excellent linear correlation was obtained (Fig. 19b). The quantity plotted along the abscissa is based on Greenwood and Williamson's formula (Ref 15) for the real area of contact, A_r, in the elastic regime:

$$\frac{A_r}{A_a p_a} = \frac{3.2}{E^* \left(\frac{\sigma_p}{R_p}\right)^{1/2}} \quad \textbf{(Eq 10)}$$

where A_a is the apparent area of contact, p_a is the apparent pressure, σ_p is the standard deviation of the composite peak-height distribution of the contacting surfaces, R_p is the composite peak curvature of the contacting surfaces, and E^* is a composite modulus that is a function of the Young's modulus and Poisson's ratio of the contacting material. The linear relationship obtained by Bhushan *et al.* (Ref 59, 60) was duplicated by Miyoshi *et al.* (Ref 61) for the same six magnetic tapes sliding on a nickel-zinc ferrite pin in a pin-on-flat experiment.

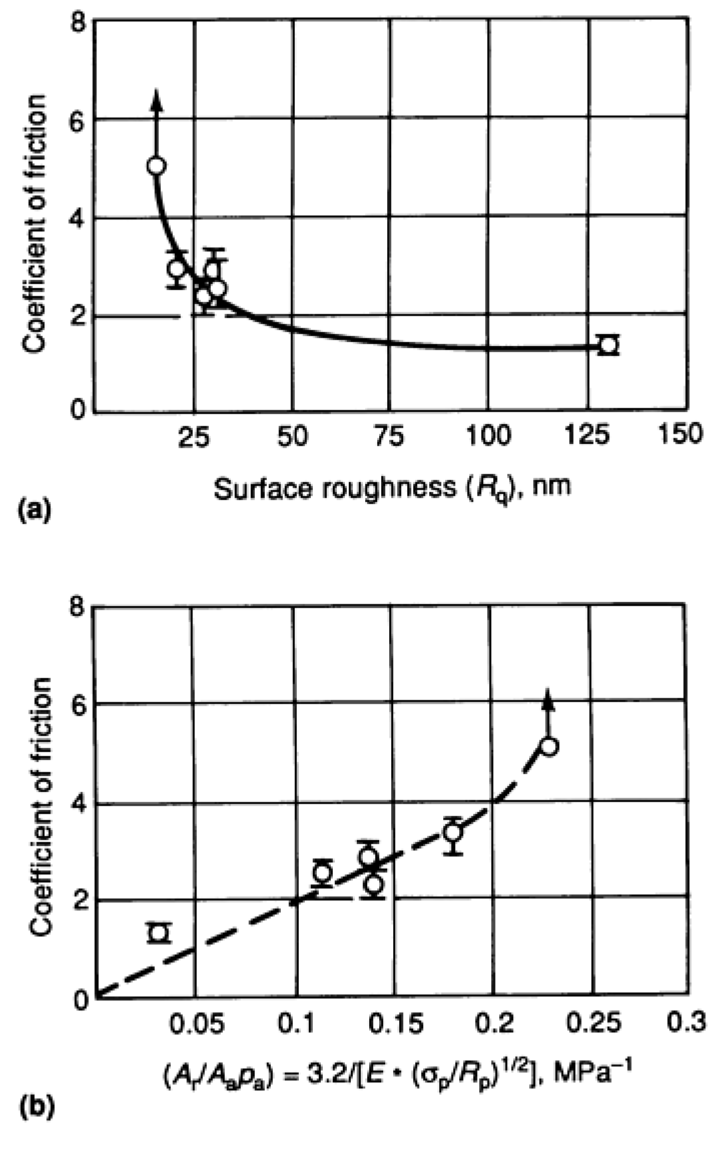
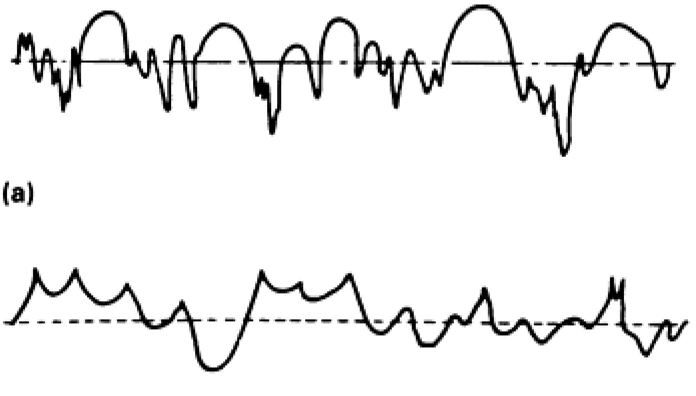


Fig. 19 Coefficient of friction for six CrO_2 magnetic tapes as a function of two parameters. (a) Coefficient of friction versus rms roughness, R_q. (b) Coefficient of friction versus the real area of contact, A_r (normalized to contact load). Source: Ref 59, 60

Lip Seals. Thomas *et al.* (Ref 62, 63) used pattern recognition techniques to correlate surface texture and lip sealing performance. They measured surface profiles of a set of rubber lip seals, some good and some leaky, and calculated a

number of surface parameters from the profiles, such as R_a, R_{sk}, and peak curvature. The groups of parameters for the good and bad seals were then separated by pattern recognition techniques. From these results, they constructed model profiles for successful and leaky sealing surfaces (Fig. 20). Although this approach is highly empirical, it can lead to a sound understanding of surface function by enabling the engineer to focus on the most probable parameters affecting performance.



(a)

(b)

Fig. 20 Reconstructions from pattern recognition analysis of profiles of the contacting surface of lip seals. (a) Ideally good seal. (b) Ideally bad seal. Source: Ref 63

Wherever possible, engineering surfaces should be assessed by evaluating those surface parameters that strongly correlate with the component function. The type and control values of these functional parameters can be determined by controlled experiments. This example of lip seals again highlights the importance of surface texture design (Ref 31). The functional performance of engineering surfaces can be optimized in a comprehensive way by proper design of their surface texture, specification of the material and manufacturing process, and development of quality control procedures.

References

1. Elements of this article have been presented in T.V. Vorburger and G.G. Hembree, "Characterization of Surface Topography," Navy Metrology R&D Program Conference Report, U.S. Department of the Navy (Corona, CA), April 1989, and in Ref 2
2. T.V. Vorburger and J. Raja, "Surface Finish Metrology Tutorial," NISTIR 89-4088, National Institute of Standards and Technology, 1990
3. T.R. Thomas, Ed., *Rough Surfaces,* Longman, London, 1982, p 189
4. "Surface Texture (Surface Roughness, Waviness, and Lay)," ANSI/ASME B46.1-1985, American Society of Mechanical Engineers, 1985
5. "Instruments for the Measurement of Surface Roughness by the Profile Method--Vocabulary," ISO 1879/1981, International Organization for Standardization, 1981
6. D.J. Whitehouse, *Characterization of Solid Surfaces,* P.E. Kane and G.R. Larrabee, Ed., Plenum Press, 1975, p 49-73
7. J.M. Bennett and J.H. Dancy, Stylus Profiling Instrument for Measuring Statistical Properties of Smooth Optical Surfaces, *Appl. Opt.,* Vol 20, 1981, p 1785
8. B. Scheffer and C. Thurel, Données de base de la Realisation d'un Calculateur R et W, *Méc. Matér. Electr.,* Vol 286, 1973, p 19
9. J. Bielle, Functional Needs, Machining Conditions, and Economics of Surface Finishing, *Prec. Eng.,* Vol 7, 1985, p 31
10. D.J. Whitehouse, The Parameter Rash--Is There a Cure?, *Wear,* Vol 83, 1982, p 75
11. J.M. Bennett and L. Mattsson, *Introduction to Surface Roughness and Scattering,* Optical Society of

America, 1989

12. P.K. Hansma, V.B. Elings, O. Marti, and C.E. Bracker, Scanning Tunneling Microscopy and Atomic Force Microscopy: Application to Biology and Technology, *Science,* Vol 242, 1988, p 209
13. R.D. Young, "The National Measurement System for Surface Finish," NBSIR 75-927, National Bureau of Standards, 1976
14. "Surface Roughness--Terminology--Part 1: Surface and Its Parameters." ISO 4287/1, International Organization for Standardization, 1984
15. J.A. Greenwood and J.B.P. Williamson, Contact of Nominally Flat Surfaces, *Proc. R. Soc. (London) A,* Vol A295, 1966, p 300
16. A.W. Bush and R.D. Gibson, A Theoretical Investigation of Thermal Contact Conductance, *Appl. Energy,* Vol 5, 1979, p 11
17. P.R. Nayak, Random Process Model of Rough Surfaces, *J. Lubr. Technol. (Trans. ASME),* Vol 93, 1971, p 398
18. J.S. Bendat and A.G. Piersol, *Random Data: Analysis and Measurement Procedures,* Wiley-Interscience, 1971
19. M.S. Longuet-Higgins, The Statistical Analysis of a Random, Moving Surface, *Trans. R. Soc. (London) A,* Vol 249A, 1957, p 321
20. R.B. Blackman and J.W. Tukey, *The Measurement of Power Spectra,* Dover, 1959
21. R.S. Sayles, *Rough Surfaces,* T.R. Thomas, Ed., Longman, London, 1982, chap 5
22. J.B.P. Williamson, *Rough Surfaces,* T.R. Thomas, Ed., Longman, London, 1982, chap 1
23. J.M. Elson and J.M. Bennett, Relation Between the Angular Dependence of Scattering and the Statistical Properties of Optical Surfaces, *J. Opt. Soc. Am.,* Vol 69, 1979, p 31
24. E.O. Brigham, *The Fast Fourier Transform,* Prentice-Hall, 1974, chap 10
25. T.V. Vorburger, "FASTMENU: A Set of FORTRAN Programs for Analyzing Surface Texture," NBSIR 83-2703, National Bureau of Standards, 1983, chap 11
26. R.L. McKenzie, Ed., "NIST Standard Reference Materials Catalog 1990-1991," National Institute of Standards and Technology, 1990, p 124
27. E.C. Teague, F.E. Scire, and T.V. Vorburger, Sinusoidal Profile Precision Roughness Specimens, *Wear,* Vol 83, 1982, p 61
28. E.L. Church, T.V. Vorburger, and J.C. Wyant, Direct Comparison of Mechanical and Optical Measurements of the Finish of Precision Machined Optical Surfaces, *Opt. Eng.,* Vol 24, 1985, p 388
29. K.J. Stout and E.J. Davis, Surface Topography of Cylinder Bores--Relationship Between Manufacture, Characterization, and Function, *Wear,* Vol 95, 1984, p 111
30. "Measurement of Surface Roughness, Parameter R_k, R_{pk}, R_{vk}, M_{r1}, M_{r2} for the Description of the Material Portion (Profile Bearing Length Ratio) in the Roughness Profile," DIN 4776-1985, Deutsches Institut für Normung, 1985
31. J.F. Song and T.V. Vorburger, Standard Reference Specimens in Quality Control of Engineering Surfaces, *J. Res. Natl. Inst. Stand. Technol.,* Vol 96, 1991, p 271
32. D.W. Blakely and G.A. Somorjai, The Dehydrogenation and Hydrogenolysis of Cyclohexane and Cyclohexene on Stepped (High Miller Index) Platinum Surfaces, *J. Catal.,* Vol 42, 1976, p 181
33. R.E. Reason, *Modern Workshop Technology,* Vol 2, *Processes,* 3rd ed., H.W. Baker, Ed., Macmillan, 1970, chap 23
34. J.D. Garratt, New Stylus Instrument With a Wide Dynamic Range for Use in Surface Metrology, *Prec. Eng.,* Vol 4, 1982, p 145
35. R.E. Reason, Surface Finish and Its Measurement, *J. Inst. Prod. Eng.,* Vol 23, 1944, p 347
36. J.B. Bryan, The Abbé Principle Revisited--An Updated Interpretation, *Prec. Eng.,* Vol 1, 1979, p 129
37. E.C. Teague, R.D. Young, F.E. Scire, and D. Gilsinn, Para-Flex Stage for Microtopographic Mapping, *Rev. Sci. Instrum.,* Vol 59, 1988, p 67

38. T.V. Vorburger, E.C. Teague, F.E. Scire, and F.W. Rosberry, Measurements of Stylus Radii, *Wear,* Vol 57, 1979, p 39
39. J.F. Song and T.V. Vorburger, Stylus Profiling at High Resolution and Low Force, *Appl. Opt.,* Vol 30, 1991, p 42
40. J.F. Song, Random Profile Precision Roughness Calibration Specimens, *Surf. Topog.,* Vol 1, 1988, p 303
41. Calibration Specimens--Stylus Instruments--Types, Calibration and Use of Specimens, ISO 5436-1985, International Organization for Standardization, 1985
42. T.R. Thomas, Ed., *Rough Surfaces,* Longman, London, 1982, p 24-25
43. M.N.H. Damir, Error in Measurement Due to Stylus Kinematics, *Wear,* Vol 26, 1973, p 219
44. S. Ajioka, The Dynamic Response of Stylus, *Bull. Jpn. Soc. Prec. Eng.,* Vol 1, 1966, p 228
45. J.I. McCool, Assessing the Effect of Stylus Tip Radius and Flight on Surface Topography Measurements, *J. Tribology (Trans. ASME),* Vol 202, 1984
46. E.C. Teague, "Evaluation, Revision, and Application of the NBS Stylus/Computer System for the Measurement of Surface Roughness," Tech. Note 902, National Bureau of Standards, 1976
47. J.C. Wyant, C.L. Koliopoulos, B. Bhushan, and O.E. George, An Optical Profilometer for Surface Characterization of Magnetic Media, *ASLE Trans.,* Vol 27, 1984, p 101; and B. Bhushan, J.C. Wyant, and C.L. Koliopoulos, Measurement of Surface Topography of Magnetic Tapes by Mirau Interferometry, *Appl. Opt.,* Vol 24, 1985, p 1489
48. T.V. Vorburger, "Appendix A: Measurement Conditions and Sources of Uncertainty for NIST Roughness and Step Height Calibration Reports," unpublished
49. J. Hasing, Herstellung und Eigenschaften von Referenznormalen für das Einstellen von Oberflachenmessgeraten, *Werkstattstechnik,* Vol 55, 1965, p 380
50. J.F. Song, T.V. Vorburger, and P. Rubert, Comparison Between Precision Roughness Master Specimens and Their Electroformed Replicas, *Prec. Eng.,* Vol 14, 1992, p 84
51. B.J. Griffiths, Manufacturing Surface Design and Monitoring for Performance, *Surf. Topog.,* Vol 1, 1988, p 61
52. H.K. Tonshoff and E. Brinksmeier, Determination of the Mechanical and Thermal Influences on Machined Surfaces by Microhardness and Residual Stress Analysis, *CIRP Ann.,* Vol 29 (No. 2), 1980, p 519
53. P.A. Willermet and S.K. Kandah, Wear Asymmetry--A Comparison of the Wear Volumes of the Rotating and Stationary Balls in the Four-Ball Machine, *ASLE Trans.,* Vol 26, 1982, p 173
54. I.M. Feng, A New Approach in Interpreting the Four-Ball Wear Results, *Wear,* Vol 5, 1962, p 275
55. E.P. Whitenton and P.J. Blau, A Comparison of Methods for Determining Wear Volumes and Surface Parameters of Spherically Tipped Sliders, *Wear,* Vol 124, 1988, p 291
56. E.P. Whitenton and D.E. Deckman, Measuring Matching Wear Scars on Balls and Flats, *Surf. Topog.,* Vol 2, 1989, p 205
57. C.H. Bovington, Surface Finish and Engine Testing of Lubricants: An Industrialist View, *Surf. Topog.,* Vol 1, 1988, p 483
58. E.J. Davis, P.J. Sullivan, and K.J. Stout, The Application of 3-D Topography to Engine Bore Surfaces, *Surf. Topog.,* Vol 1, 1988, p 63; K.J. Stout, E.J. Davis, and P.J. Sullivan, *Atlas of Machined Surfaces,* Chapman and Hall, London, 1990
59. B. Bhushan, *Tribology and Mechanics of Magnetic Storage Devices,* Springer-Verlag, 1990
60. B. Bhushan, R.L. Bradshaw, and B.S. Sharma, Friction in Magnetic Tapes II: Role of Physical Properties, *ASLE Trans.,* Vol 27, 1984, p 89
61. K. Miyoshi, D.H. Buckley, and B. Bhushan, "Friction and Morphology of Magnetic Tapes in Sliding Contact With Nickel-Zinc Ferrite," Technical Paper 2267, National Aeronautics and Space Administration, 1984
62. T.R. Thomas, C.F. Holmes, H.T. McAdams, and J.C. Bernard, Surface Microgeometry of Lip Seals Related to Their Performance, Paper J2, *Proc. 7th Int. Conf. on Fluid Sealing,* BHRA Fluid Engineering, Cranfield UK, 1975

63. T.R. Thomas and R.S. Sayles, *Rough Surfaces,* T.R. Thomas, Ed., Longman, London, 1982, p 231-233

Surface Topography and Image Analysis (Area)

Eric P. Whitenton, National Institute of Standards and Technology

Introduction

SEVERAL CONCEPTS and methods involved in the topography and image analysis of engineered and worn surfaces are described in this article, in terms of the past, present, and future of this characterization technique. Although linear profilometry has long been used in materials research to study both machined and worn surfaces (Ref 1), there is typically more information about a surface in a scanned area profile (Ref 2). The fact that both computers and machines that can perform scanned topographical profile measurements over an entire area are becoming more powerful and less costly, combined with the scanned area profile advantage, represent two of several reasons why the techniques for analyzing profile information are becoming similar to techniques used for analyzing optical and scanning electron microscope (SEM) images in materials research.

Image analysis of optical and SEM photomicrographs have been used for many years for various purposes related to materials science (Ref 3, 4, 5). Both optical and SEM images are essentially two-dimensional *x*,*y* arrays of numerical values. Each value represents the intensity of the image at that *x*,*y* location. Generally, area profiling machines also produce an *x*,*y* array, but each value represents a *z* height at that location. If intensity and *z* height are allowed to be interchangeable, where one can be substituted for the other, then the same equipment, techniques, and computer software can be used to analyze both. This simplifies the data analysis tasks of the researcher by unifying many of the techniques that must be learned. One machine that applies this approach uses much of the same hardware and software to interchangeably perform laser scanning tomography, infrared (IR) transmission photomicroscopy, and noncontact optical profilometry (Ref 6).

Historically, the topographical analysis of machined surfaces has predominantly consisted of compiling statistics of geometrical properties, such as average slope or the root mean square (rms) of the *z* heights. The theory behind this is described in detail in the literature (Ref 1). Techniques like this have been of limited use in the characterization of worn surfaces, particularly those that are severely worn, but can be efficiently performed in an image analysis environment. Examples are given in this article.

Image analysis is also becoming increasingly useful to pick out, characterize, manipulate, and classify the features on a surface individually, as well as in groups. It seems unlikely that purely statistical techniques will ever reach this level of sophistication. Investigators may soon see surfaces described in terms of the organizational structure of features, instead of rms. This article discusses a few of the potential pitfalls, capabilities, and opportunities of this evolving tool.

A novel example of how image analysis and profiling are interrelated is in the measurement of pigment agglomeration in rubber (Ref 7). The standard procedure is to microtome the frozen rubber and examine it under an optical microscope. Using image analysis techniques, the darker-colored agglomerates are differentiated from the lighter-colored rubber, and the dispersion is computed. The researchers noticed that a stylus profile tracing of the rubber, sliced with a knife blade at room temperature, essentially yields a flat plane that has distinct holes and bumps. This is because the soft rubber "cuts" in a flat plane, whereas the harder agglomerates are not cut and protrude through the cutting plane. The number of peaks per unit area, a method long used in both image and profile analysis, is used to compute the dispersion. This method was judged to be very accurate and fast.

Definitions and Conventions. Where possible, cited reference works were selected because they present techniques in "cookbook" form. It is hoped that this encourages readers to try such techniques on their own systems.

A *topographic image* refers to an image where each *x,y* location represents a *z* height. This image is generally acquired by a scanning profiling machine. An *intensity image* refers to an image where each *x,y* location represents an intensity, and is normally obtained by SEM or video camera. A *binary image* is derived from either a topographic or an intensity image. Each *x,y* location has a value of either "0" or "1," indicating which locations in the original image have some property, such as *z* height above a threshold value or the edge of a feature as determined by local slopes. Some of the techniques discussed in this article are performed on binary images, which are described more fully in the section "Computing Differences Between Two Traces or Surfaces" and portrayed in Fig. 5. The word *image,* by itself, is intended to be very generic. It can refer to a topographic image, an intensity image, and, in certain circumstances, individual traces. A single trace is, in fact, the special case of an image with only one row of data. Note that what makes a topographic image different from an intensity image is simply the meaning of the value at each *x* and *y,* and not how it is displayed, or *rendered.* If an isometric line drawing of an intensity image is displayed, the image is still an intensity image, even though it "looks" as though it were a topographic surface. It should be remembered that all images are *single-valued functions,* which is to say that for any given *x* and *y* value, there is one and only one *z* value. The ramifications of this are discussed throughout this article.

*Motifs*were the first profile analysis technique developed especially for use on computers (Ref 8). Using a set of four simple and easily understood rules, a complex trace can be reduced to a simpler one. This technique has been used in the French automotive industry for many years, and numerous practical uses have been found (Ref 8, 9, 10, 11). Currently, these rules only apply to a two-dimensional trace. If appropriate rules were discovered, this technique could also be performed on three-dimensional images.

Surfaces are sometimes referred to as either deterministic, nondeterministic, or partially deterministic. A *deterministic surface* is a surface in which the *z* heights can be predicted if position on the surface is known. Sinusoidal (Ref 12) and step-height calibration blocks are examples. A *nondeterministic surface* has random *z* heights, such as a sand-blasted surface. Some surfaces have both a deterministic and a nondeterministic character. A ground surface often has a distinct, somewhat predictable, lay pattern with a random fine roughness superimposed on it. Such a surface is often termed *partially deterministic.*

Leveling refers to the process of defining $z = 0$ for an image. For example, a single-profile trace is taken across a flat specimen. If one side of the specimen were higher than the other side, then the trace could be leveled by subtracting a line from that trace. For an engineered surface, the line would typically be determined by performing the least-squares fit of a line to all of the data in the trace. For a worn surface, where part of the trace includes the worn area and part includes the unworn area, only some of the data in the trace would be used to determine the least squares line. The data in the unworn area only would be used to determine the least-squares line when the worn volume, or wear scar depth, was to be determined.

Implementation on Personal Computers and Data Bases. Both software (Ref 13, 14) and books (Ref 15, 16, 17, 18) have become readily available to perform image analysis on personal computers. At least one source (Ref 18) not only describes many of the techniques, but also includes software. If a profiling or other image-producing machine, such as a microscope, were under heavy use, then users could take a floppy disk containing the stored images to another work station and free the measuring equipment for others to use. Some data base programs allow images to be stored along with other textual and numeric information (Ref 19). It is also possible to have the images themselves as part of the querying process, where a user "enters" an image and the computer finds similar images (Ref 20). Thus, both the topography, or topographic image, and visual appearance, or intensity image, of a surface can be an integral part of a data base.

Point Spacing and Image Compression

The issue of how many *x,y* points to acquire in an image generally involves a compromise. If too few points are used, then valuable information can be lost. It has been shown, for example, that a surface with an exponential correlation function appears as a Gaussian correlation, unless there are at least ten data values per correlation length (Ref 21). The determination of even a simple parameter, such as rms roughness, is also affected (Ref 22, 23). When too many points are used, more mass storage and computing time per image are required than necessary. Also, the determination of noise-sensitive parameters can be adversely affected (Ref 24). This is because extremely fine point spacings may enhance the ability of the computer to record the noise in the profiling system, along with the topographic information.

One solution is to acquire as many points as possible and later discard the redundant or unimportant values. There are a variety of image data-compression techniques that remove redundant or unimportant information when the image is stored in memory or disk. The best compression technique depends on which aspects of the image are redundant or not

important to image quality. Several data-compression techniques have been proposed for surfaces of materials. One technique uses Fourier transforms (Ref 25, 26). By storing only the "important" frequencies, the amount of data can be reduced. The selection of which frequencies are not stored implies that features of that lateral size range can either be extremely small in vertical height, compared to other features, or are unimportant. Other procedures attempt to determine the "optimum" point spacing using autocorrelation functions (Ref 27), bandwidths (Ref 24), or information content (Ref 28). If variable point spacings are allowed, then motifs provide another technique (Ref 8). Many of the possible data-compression techniques do not appear to have been tried on images of surfaces of materials.

Walsh or Hadamard transforms, where a surface is modeled as a series of rectangular waves, can be used in place of Fourier transforms. This often results in less noise in the reconstructed image, although Fourier transforms may better reproduce the original peak shape (Ref 26). Although there do not appear to be any references in the literature on usage as a data-compression technique specifically for the surfaces of materials, the coefficients have been used to characterize these surfaces (Ref 29, 30). Many other data-compression techniques are also available.

Potential Pitfalls

Many of the potential pitfalls in intensity image processing are potential pitfalls in topographic image processing as well. For example, when determining the roundness of an object, the number computed is dependent on the magnification used (Ref 31). A computed area or length also depends on the scale used, this being one of the basic concepts behind fractals, which are discussed in detail in the section "Fractals, Trees, and Future Investigations" in this article.

Another pitfall is the fact that the surface is being modeled as a single-valued function in *x* and *y,* when it may in fact not be. One example is a case where a "chip" of material is curled over the side of a machined groove. There are at least three *z* heights: the top side of the curled chip, the underside of the curled chip, and the top surface of the bulk material below that chip. A profiling machine would report only the top side of the curled chip as the *z* height at that *x,y* location. Any estimate of volume would obviously be larger than the actual volume of material. Thus, an image of a surface is actually made up of only the highest points on the surface. A top view is the only truly accurate rendering of the image; other renderings, such as isometric or side views, are only approximations. This is because these other renderings give the appearance of "knowing" what is below those highest points.

An analogous situation in intensity images is the "automatic tilt correction" on some SEMs (Ref 31). Suppose an intensity image of a sphere on a steeply sloped plane is acquired and that slope is removed in software so as to make the plane appear horizontal. A side view of this situation is shown in Fig. 1. When the software attempts to "level" the image, the radius of the sphere will be elongated in the direction of the tilt and remain constant in the orthogonal direction. The sphere will then appear as an ellipsoid, and not as a sphere.

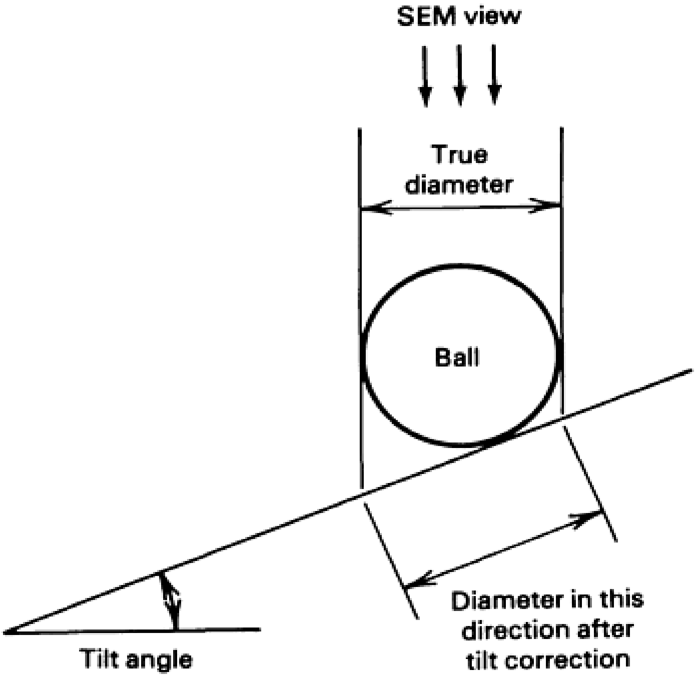


Fig. 1 Side view of a sphere on a sloped plane

Estimation and Combination of Intensity and Topographic Images

Simply displaying a topographic image as though it were an intensity image (which can be a very powerful tool) does not show the user how the surface would actually appear under a microscope. The heights are known, but the color, reflectivity, and translucency of the surface are not. Conversely, a microscope image gives clues as to the surface heights, but does not do so quantitatively. It may be obvious that a surface is pitted, for example, but the depth of those pits are not known. Three issues are therefore addressed: (1) The manipulation of an optical or SEM image to yield topographic information; (2) The rendering of topographic information that actually looks like the surface; (3) The combination of optical and topographic information together onto one rendering.

Transforming an intensity image to a topographic image can be approached in several ways. All approaches involve a "nicely behaved" characteristic of the surface. One approach matches stereo pairs. Each feature in a left-eye image is matched to the same feature in a right-eye image. When the two images are compared, the amount of lateral displacement of each feature is related to its z height. Thus, a z height image can be created. The features must be distinct and well defined for this approach to work well. An example of this in use is in the measuring of integrated circuit patterns (Ref 32).

Another approach assumes that the optical properties of the surface are relatively constant. If the original surface does not have this property, then a replica can be made and examined, instead. When properly lighted, each gray level in the intensity image is proportional to the slope of the surface at that location (Ref 33). The topographic image can therefore be found by integrating the intensity image.

An example of a third approach is a wear scar on a ball. The volumes of such scars are often determined by measuring the scar width in an intensity image and assuming that the scar is relatively flat or of a fixed radius in z (Ref 34). However, the scars may be of unknown or varying radii. More accurate volume estimates can be obtained by outlining the edge of the worn scar and assuming the outlines are connected by lines or curves across that scar (Ref 35). This is shown in Fig. 2, where the surface has, in effect, been estimated from its intensity image and the known geometries in that image.

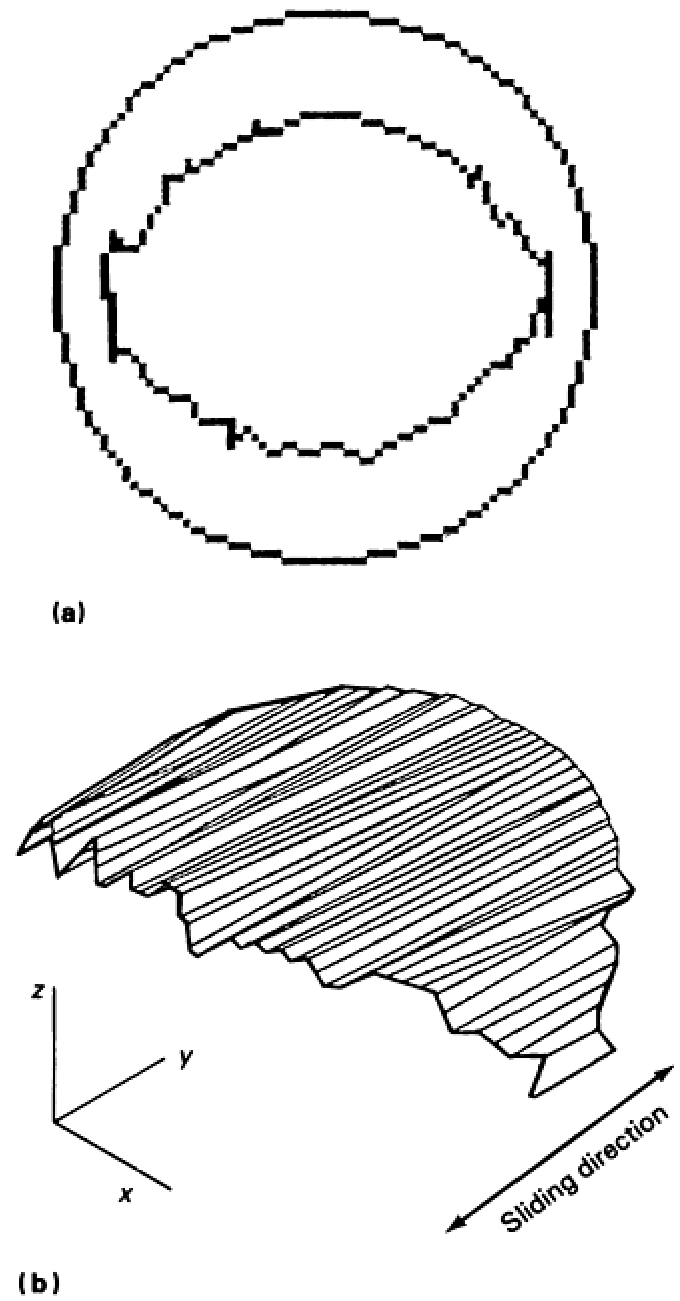


Fig. 2 Example of estimating a topographic image from an intensity image using known geometries

A nonrotating ball was slid repeatedly against abrasive paper in the *y* direction, forming a scar on the ball. An optical photomicrograph that looks down onto the scar was taken, digitized, and the intensity image was shown on the computer screen. The user then traced the outline of the scar using a pointing device. This is shown as the near-elliptical shape in Fig. 2(a). The software then assumed that the *x,y* location of the center of the scar coincided with the *x,y* coordinate of the center of the ball. Knowing the radius of the ball, the software then computed the *z* heights of all the *x,y* points on the outline of the scar, because they must lie on the sphere. To estimate the *z* values inside the scar outline, the values of the outline were connected by straight lines in the *y* direction, as shown in Fig. 2(b).

Rendering and Combining Images. Actually transforming a topographic image to an intensity image is rarely done for surfaces of materials. The appearance of a surface under a microscope is typically approximated by simply rendering the topographic image as an isometric view. Isometric views can be generated by most image analysis software. The simplest isometric view is a stick-figure type of drawing, where no attempt is made to show how a light source would interact with the surface (Ref 18). These views may or may not have hidden lines removed. The next level of sophistication assumes that the optical properties are constant across the entire surface. One or more light sources are

assigned locations in space, and the view is "shaded," giving a more realistic appearance. Some software takes into account the shadows that one feature casts onto another, whereas others do not. Often, however, the optical properties of real surfaces are not constant across the entire surface.

Given optical properties maps of reflectivity, for example, some software can create very realistic renderings (Ref 36). An intensity image of a properly lighted surface can be used as a reflectivity map. Therefore, such software can be used to combine an intensity image and a topographic image of the same area to produce a rendering that exhibits both optical and topographic qualities of the surface.

Relating Two- and Three-Dimensional Parameters

Situations in which researchers have preferred the more traditional two-dimensional parameters have occurred. One example is the case where a large body of two-dimensional data has already been collected and there is a need to compare newly acquired data with previously obtained values. Even in these cases, the ability to select which two-dimensional trace to use for analysis from a three-dimensional topographic image is sometimes necessary (Ref 37). Additionally, the repetitive application of the analysis for a large number of traces can provide statistical information as to the repeatability of the results obtained for a given specimen (Ref 38, 39, 40, 41, 42). When applied to worn surfaces, a two-dimensional parameter can often be plotted as a function of sliding distance, giving clues as to the mechanisms involved (Ref 43). It is possible to estimate three-dimensional parameters from two orthogonal traces. This has been applied to mold surface finish (Ref 44) and has been used in the comparison of the fractal dimension (discussed later in this article) both with and across the lay of engineered surfaces (Ref 45).

However, better results are often obtained from full images (Ref 46). Many of the customary two-dimensional parameters are easily extendable to three dimensions. Perhaps the best-studied parameters in both two and three dimensions are roughness parameters, such as rms values. Generally, two-dimensional roughness parameters have smaller values than their three-dimensional counterparts for nondeterministic surfaces, and have about equal values for deterministic surfaces. This result is derived from both theoretical work (Ref 1) and actual data (Ref 38).

There are two explanations for this result. One is that single traces have a high probability of missing the highest peaks on a surface, whereas an area profile has a much better chance of taking these into account (Ref 1). Another explanation involves the fact that nondeterministic surfaces have waviness in both the x and y directions (Ref 47). Waviness in the x direction is generally removed by filtering for both the two- and three-dimensional roughness calculations. The two-dimensional calculation always removes waviness in the y direction, because each trace is leveled individually. The three-dimensional calculation, where the same plane is subtracted from all of the trees, does not do so unless a filter is specifically applied to the image in the y direction. Thus, the three-dimensional roughness parameter may or may not include the waviness in the y direction, depending on how the parameter is computed.

When analyzing worn surfaces, some area profiling machines use the unworn part of a surface as a reference. This is done by fitting the unworn part of each trace to a line, and subtracting the line from that trace (Ref 41, 43). An example of this is shown in Fig. 3. Typically, this is performed because of drift problems while the traces are being acquired and to make the worn volume measurements more accurate. The effect is to filter the waviness in the y direction. One might therefore expect that a three-dimensional roughness parameter computed from this image would be more nearly equal to the two-dimensional equivalent than the same parameter applied to an image acquired by a machine that only uses its own reference plane. However, this does not appear to have been rigorously demonstrated.

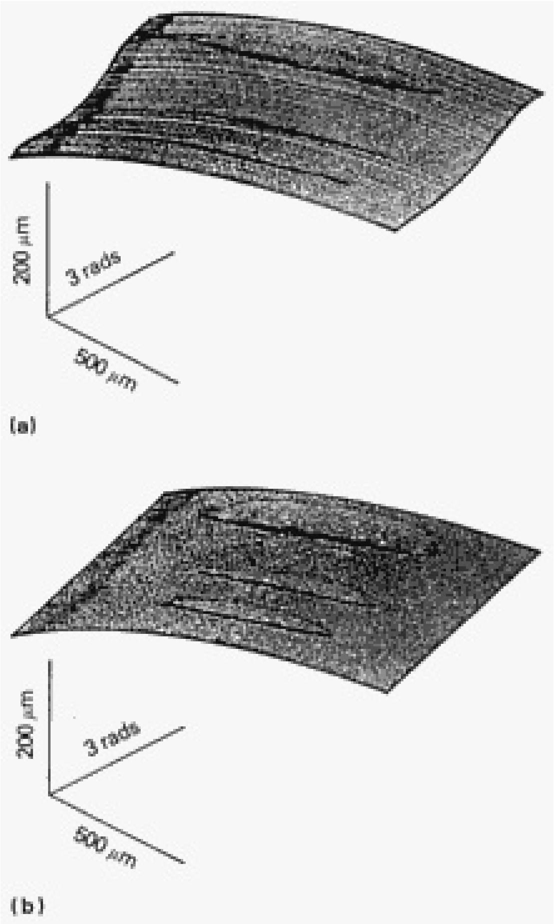


Fig. 3 An x,θ,z coordinate image of the doughnut-shaped scar on the top ball in a four-ball test

Figure 3(a) shows the "as traced" data. Note the vertical undulation of the surface. This is due primarily to mechanical errors in the motor stage used to hold the ball during image acquisition. For each trace, the unworn area can be fit to a line, and that line used to make the trace level with respect to the other traces. This is shown in Fig. 3(b).

The relationships between the two- and three-dimensional values for other parameters are not as well documented as roughness. Other statistical parameters, such as skewness and kurtosis (which help characterize the distribution of *z* heights), have been computed for both engineered (Ref 42, 46) and worn (Ref 48) surfaces. Aspect ratio parameters have been proposed for circular wear scars (Ref 40) and for the features in worn areas (Ref 43). Fractal dimensions can also be determined in three dimensions (Ref 49, 50). It should be remembered that the values obtained for many two-dimensional parameters are often quite different, depending on the direction of the trace. Rms roughness (Ref 51), autocorrelation (Ref 52), and fractal dimension (Ref 45) are examples of this.

Lessons from Two-Dimensional Analysis

Example 1: Understanding How a Parameter Behaves.

In the late 1970s, it was discovered that there is nearly the same linear relationship between the log of the wavelength and the log of the normalized power spectral density for a very large variety of surfaces (Ref 53). These surfaces span almost nine orders of magnitude in size. Values for motorways, concrete, grass runways, lava-flows, ship hulls, honed raceways, ground disks, ring-lapped balls, and other surfaces were used. An amazingly universal characteristic of real surfaces was discovered. Today, it is known that this occurs because these surfaces are fractal in nature (Ref 45). Imagine that a researcher does not know of this universality, but notices that this relationship exists for a particular set of surfaces. It might be tempting to assume that something was unique about these particular surfaces, when, in actuality, certain parameters behave in certain ways regardless of the type of surface.

Example 2: Determining a Reference Line or Parameter Value on a Pitted or Grooved Surface.

Certain features on the surfaces of some materials do not affect performance and should be ignored when leveling, fitting, or determining roughness parameters. An application where a small roughness is required on a surface, except for periodic deep scratches to contain lubricant, is one example. The porosity in many ceramics is another.

One approach to evaluating these types of surfaces is to be able to selectively ignore certain *z* values, based on an appropriate criterion. One example of this is to ignore *z* values that are several standard deviations away from the average (Ref 54). Wide scratches can be detected and ignored by looking for clusterings of these outliers.

It should be noted that a single trace cannot distinguish between a scratch and a pit. In some applications, such as the characterization of corrosive pitting, that information may be desirable. Image analysis can determine such differences in several ways, such as by computing aspect ratio parameters and by pattern matching.

Example 3: Designing Parameters.

When two-dimensional parameters became commonly used in materials research, a proliferation of many similar, but not identical, parameters appeared in the literature. One study used correlation analysis to examine 30 parameters applied to various engineered surfaces (Ref 55). Many of these parameters were found to be highly correlated, and several were selected as being the least redundant. It was suggested that all or some subset of these few should be used to study engineered surfaces, because they each revealed a different characteristic of these surfaces.

Other researchers have performed similar studies using correlation (Ref 56) and cluster analysis (Ref 57). The popularization of three-dimensional parameters may, in some ways, worsen the proliferation of parameters. However, image analysis can be thought of as either a language or tool box of techniques for optimizing parameters to suit particular needs. Evaluation procedures can be custom built from combinations of relatively standard image operations.

The idea of designing a parameter for an application has found its way into two-dimensional parameters. Examples include the German standard DIN 4776 (R_k) (Ref 11, 58), functional filtering (Ref 1, 10), and the French standard NF05-015 (motifs) (Ref 8, 9, 10). Invariably, some combinations will prove useful in a wide range of applications, whereas others will fall into obscurity.

Selecting an Appropriate Coordinate System

Figure 4 shows a few of the worn specimen/coordinate system combinations possible. Figure 4(a) shows an x,y,z coordinate system used for the wear track on a flat in a pin-on-flat test. The left side of Fig. 4(b) shows an x,θ,z coordinate system used for the wear track on a fixed cylinder in a rotating cylinder on a fixed-cylinder wear test. The right side of Fig. 4(b) shows an $x,\theta z$ coordinate system used for the wear scar on a top ball in a four-ball test. Figure 4(c) shows a θ_1, θ_2,z coordinate system used to characterize an entire ball surface after having been used in a ball-bearing assembly.

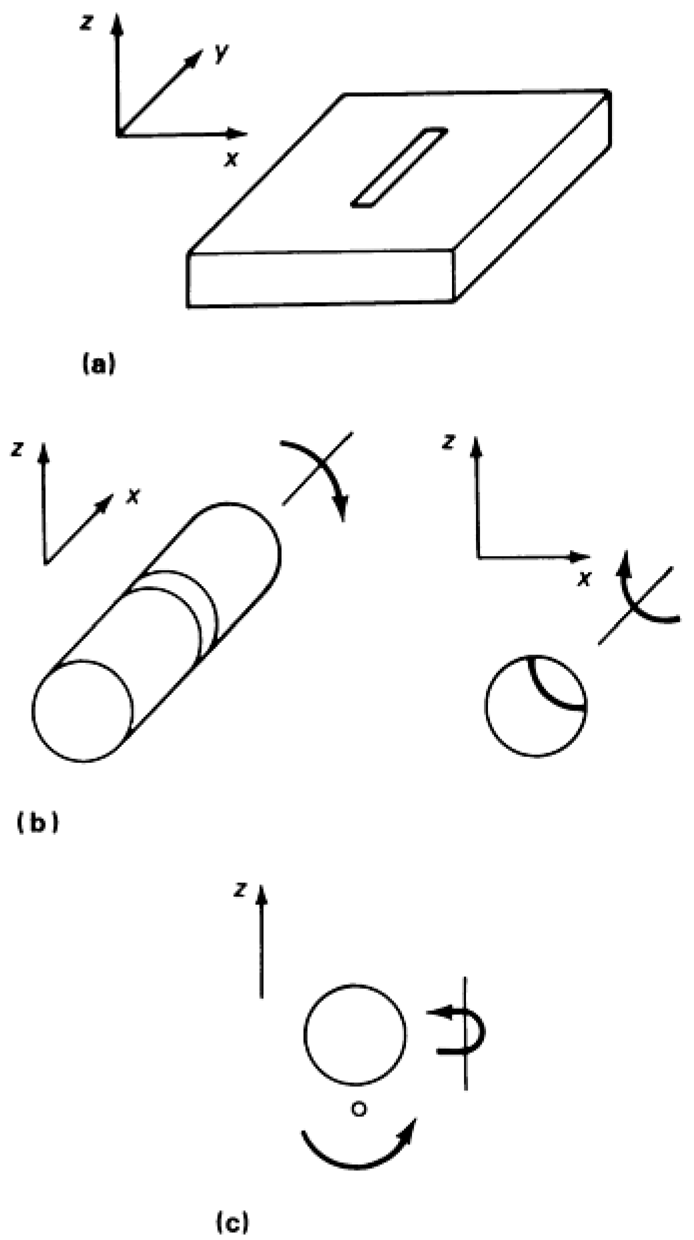


Fig. 4 Possible worn specimen/coordinate systems

The geometry of the area of interest generally determines which coordinate system is the most efficient to use. Take the example of a ball. The typical *x,y,z* coordinates can be used if the feature of interest were the wear scar on a ball in a test where the ball slides on a flat without rotating. However, an $x,\theta z$ system may be more efficient if it were the scar on the top ball in a four-ball test (Ref 40). A θ_1, θ_2,z system can be used if the entire ball surface is of interest, as in the case of ball bearings in head/disk assemblies (Ref 59) or in the evaluation of sphericity (Ref 60). Combinations of coordinate systems can be used on the same ball (Ref 61). A θ_1, θ_2,z system can be used to get an overall view of the ball, and an *x,y,z* system can be used to "zoom in" on specific features. Sometimes, $\theta_1\theta_2,z$ coordinate systems are scaled as though they were *x,y,z* coordinates (Ref 41). This can easily be done if the diameter of the ball is known. Bores and holes (Ref 62), as well as valve seats (Ref 63, 64), have been characterized in x,θ,z coordinate systems.

The x,θ,z coordinate system is sometimes referred to as a cylindrical coordinate system. However, as Fig. 4 shows, both cylinders and spherical balls can require the use of this system. The x,θ_1,θ_2 coordinate system is sometimes referred to as a spherical coordinate system. As noted above, a spherical ball can be profiled using *x,y,z* or x,θ,z coordinates, as well. Thus, these names should be used carefully. When reading the literature, for example, it is occasionally easy to confuse a cylindrical specimen with a cylindrical coordinate system.

The type of analysis to be performed can also affect the coordinate system chosen for use. For example, a planar machined surface can be traced using an x,θ,z coordinate system, where the traces radiate from some central location. When used in conjunction with autocorrelation functions, these can be used to graphically characterize the lay of a surface (Ref 52). When used in conjunction with cross-correlation functions, these can also be used to quantify the isotropy of a surface (Ref 65, 66).

Specialized hardware is generally required for the acquisition of images using alternate coordinate systems. The analysis software may need to be modified, as well. The calculation of worn volume, for example, may require a different equation for x,y,z and x,θ,z coordinate systems (Ref 40).

Computing Differences Between Two Traces or Surfaces

Perhaps the most commonly performed manipulation of topographic data, whether in the form of linear traces or images over an area, is computing the difference between two traces or images. This fact is important, because although it is one of the simplest manipulations, it is also prone to potentially large errors if not done carefully. Examples that illustrate this point and techniques for avoiding these errors are discussed below. It is important to remember, particularly in this section of the article, that the word "image" is used for both single tracings from a standard two-dimensional profiling machine and true images.

Example 4: Determining a Reference.

Often, a second image is computed from an original image and the difference between the two is derived. When leveling, for example, a reference line or plane is often fit to some or all of the image, and that line or plane is subtracted from the image. Different types of fits can be performed, and different reference lines or planes will result. Research has been conducted to compare various types of fits (Ref 67). It was found that the least-squares fit is acceptable for nearly level surfaces; orthogonal least-squares fit is better for steeply sloped surfaces; and geometric mean is preferred when the data values in the image have a log-normal distribution. The problem of ignoring outliers in the determination of a reference has been discussed above.

Example 5: Roughness, Waviness, and Error of Form.

Another example of computing a second image and deriving the difference is in the separation of an image of a machined surface into roughness, waviness, and error of form (Ref 68). Roughness consists of the finer irregularities. Waviness is the more widely spaced component of surface texture. The two components together are referred to as surface texture. Error of form is the deviation from the nominal surface not included in surface texture. These components generally result from different aspects of the machining process. An example is a ground surface. The roughness can result from the grinding wheel-workpiece interaction, the waviness from machine vibration, and error of form from errors in the guides that control the movement of the grinding wheel over the workpiece.

Roughness is often modeled as the high-frequency component, waviness as a mid-frequency component, and error of form as the lowest-frequency component of a surface. In theory, if an image of a surface was divided into these separate components, and these components were recombined, the result would be to recreate the original image. In practice, however, significant distortions often result.

Perhaps the best-known example of this is the acquisition of a roughness trace from a standard profiling device (Ref 69). Electronic filters allow the higher frequencies in the z height signal to pass through while blocking the lower frequencies. Thus, an image of the roughness component of the original image is obtained. The difference between the roughness image and the original image gives an indication of the waviness and error of form components of the surface. However, the roughness image is distorted, because of time lags in the electronic filters. The difference image of the other surface components is therefore also distorted. This effect can be minimized using modern digital filtering techniques, which do not introduce time-lag errors. Standards are currently being developed for these (Ref 70).

Example 6: Error Correction.

The differences between two images are also used to correct for errors in the z reference plane. Most profiling devices have some form of a precisely flat surface, which defines $z = 0$. Errors in this reference plane are often reproducible and

can be measured. An error image can thus be created and stored for later use. When the device is used to measure surfaces, this error image can be recalled and subtracted from the acquired topographic images to increase their accuracy (Ref 71). A similar technique can be used for intensity images to compensate for uneven illumination.

Example 7: Comparing Mated Surfaces.

Wear studies that examine the difference between two mated surfaces have been made. In one study, the differences in the roughness images of two surfaces that had been in sliding contact with each other were used to characterize the conformity between them (Ref 72). Errors associated with using a roughness image have been discussed above. Ignoring waviness when modeling the way that two surfaces interact can adversely affect the results in some situations and therefore must be done with care (Ref 73). Another issue is that of the elastic deformation of the surfaces while they were mated. The topographies of the two surfaces while they were pressed together under load is undoubtedly different from their topographies while traced. Various researchers have attempted to model this (Ref 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84).

Example 8: Determining Worn Volumes.

Described below are four areas of concern.

The Difference Image. The worn volumes of wear scars are often computed by first subtracting the image of an idealized unworn surface from the image of a worn surface. Either lines or planes can be used for a flat specimen, and circles for a ball or cylinder (Ref 40). Figures 3, 5, and 6 exemplify this. Figure 3(b) shows an image with x,θ,z coordinates of the doughnut-shaped wear scar of the top ball in a four-ball wear test. For each trace in the image, a least-squares circle is determined from unworn areas on either side of the scar, and that entire trace is then subtracted from this circle.

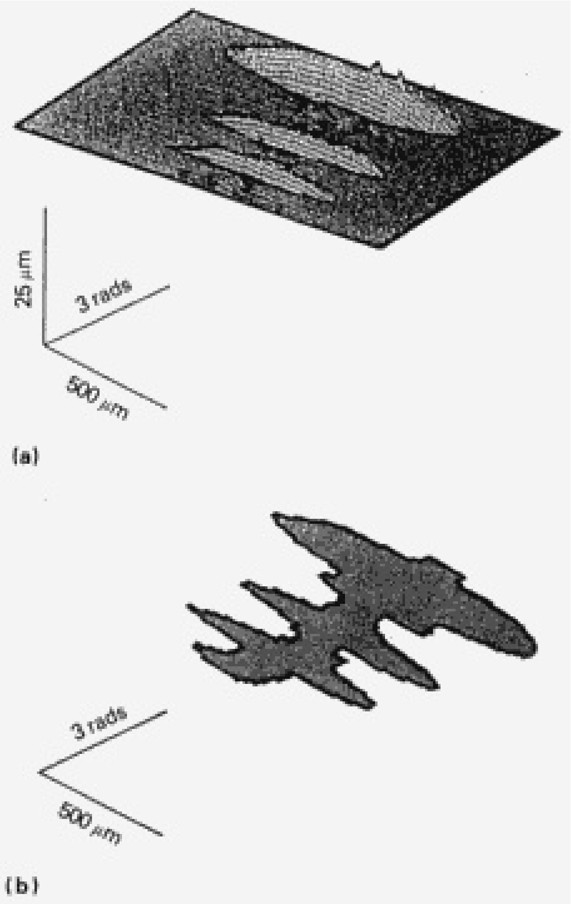


Fig. 5 (a) Difference image derived from image in Fig. 3(b). (b) Binary image

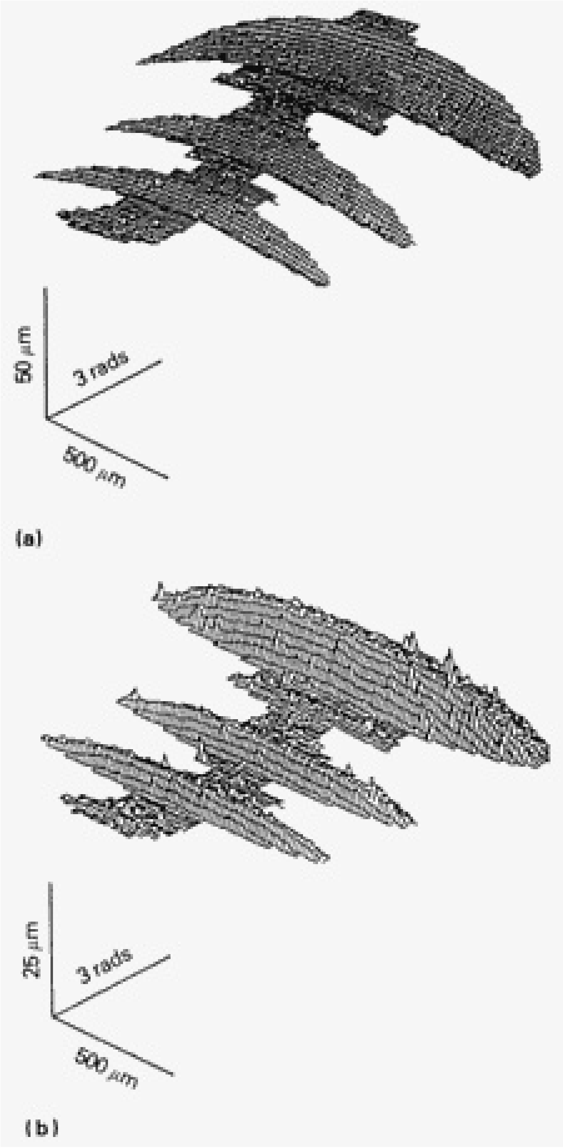


Fig. 6 (a) Worn area of image in Fig. 3(b). (b) Worn area of difference image shown in Fig. 5(a)

This new image is referred to as a difference image, and is shown in Fig. 5(a). It represents the difference between an unworn and worn ball. Where there has been a net loss of material, the difference image will have a positive value. Where there has been a net gain of material, the difference image will have a negative value. Where there has been no net change of material, the difference image will have a value close to zero. Values significantly different from zero can then be used to determine which areas of the image are worn and which are unworn.

A binary image is shown in Fig. 5(b). For each *x*,*y* location, the binary image has a value of 1 if that location is to be considered a part of the wear scar, and a value of 0, otherwise. This binary image can then be used to "eliminate" parts of the original image and difference image that are not part of the wear scar, and should therefore not be considered in any statistics computed.

Figure 6(a) shows just the worn area of the image in Fig. 3(b). Curvature, surface area, or roughness, for example, can be computed from this image. Figure 6(b) shows just the worn area of the difference image shown in Fig. 5(a). Worn volume can be computed from this image.

Alignment. The image of the unworn surface need not be idealized, but may actually have been measured before the wear test. Examples of this include the wear of copper (Ref 85), teeth (Ref 86), valve seats (Ref 63), and chemically active scuffed bearing surfaces (Ref 87, 88). The electroplating process can also be studied by comparing the topography of a surface during the various stages of plating (Ref 89). One source of error is the problem of aligning the "before" and

"after" images. Proper alignment of the worn and unworn surface images can be aided by microhardness indents (Ref 63) or other markings on the surface. If there are features on the specimen that are known to have not worn and are distinctive, then these features can be adequate substitutes for special markings (Ref 86).

Effects of Digitization. Another potential source of error when subtracting two surfaces is the fact that the worn volumes may be very small in relation to the volumes of the surfaces, especially for nonplanar specimens, such as balls. This situation results in the subtraction of large, nearly equal numbers, which is a well-known source of error in computer computations (Ref 90). After an operation, such as leveling, is performed on an image, the image should be rescaled so as to fully utilize all of the bits used to store that image. This allows subsequent operations to be performed at as high a resolution as possible, minimizing cumulative errors. The original image should be acquired and stored using as much resolution as possible. A combination of the vertical resolution of the profiling device and the number of bits actually used when the height signal is digitized, determines the useful resolution of an image.

Suppose that the noise level of a profiling device is on the order of 0.1 μm (4 μin.), with a vertical range of 1 mm (0.04 in.), represented by a voltage of 0 to 5 V. The analog-to-digital (A/D) converter acquiring the image has 12 bits over the range of 0 to 10 V. Because the A/D has a voltage range twice that of the profiling device, half of the resolution, or one bit, of the A/D will never be used. Note also that if the voltages actually digitized for this particular image range from 1.0 to 3.5 V, yet another bit has been wasted. The resolution of the A/D is about 2.5 mV, which corresponds to about 0.5 μm (20 μin.). This is a factor of five worse than the profiling device. With the appropriate electronics, the 1.0 to 3.5 V could be mapped to all 12 bits, resulting in a resolution of about 0.6 mV, or 0.1 μm (4 μin.). This more fully exploits the profiling device resolution. Of course, there are other issues that affect the useful resolution of the A/D, such as frequency response and aperture uncertainty (Ref 91).

Effects of Large Slopes and Positioning Errors. When subtracting two surfaces that contain large slopes, the result can be sensitive to lateral positioning errors. Profiling machines that acquire the topographic image while the z sensor is in motion are particularly prone to this problem, because of variations in the sensor velocity. This can be minimized by using an interferometer, linear optical encoder, or other lateral position sensor to control the data acquisition. Figure 7 shows a 10 mm (0.4 in.) diameter ball with a 2 mm (0.08 in.) wide scar.

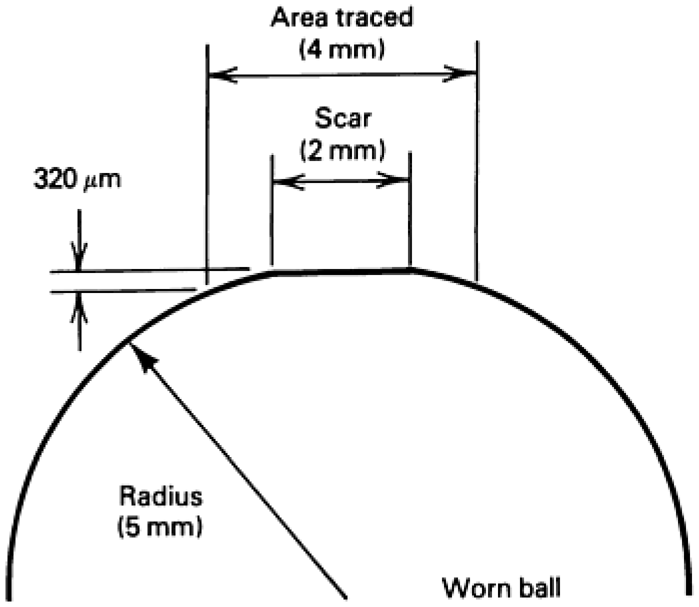


Fig. 7 Side view of a worn ball from a nonrotating ball on flat wear test, which is to be traced with a profiling device

Suppose that a 4 mm (0.16 in.) wide area is traced. This ensures that enough of the unworn portion of the ball is in the topographic image to use as a reference. The full-scale z height range would be about 320 μm (13 mils). The slope of the reference area would vary from around 11.5 to 23.6°. A lateral positioning error of 2 μm (80 μin.) therefore results in an

error of 0.4 to 0.9 μm (16 to 36 μin.) in the *z* height measurements in the reference area. This is a manageable error at a 320 μm (13 mil) full-scale height.

Suppose that the scar itself is relatively flat, with a *z* height full-scale range of 1 μm (40 μin.). When the difference between the unworn and worn ball is examined, the full-scale range of the difference image is therefore only 2 μm (80 μin.) or less. The errors in the reference area in the *z* direction are large when compared to the *z* heights of the scar itself, 20 to 45%, in this case. This situation makes it difficult to distinguish between the scar and the reference area near the edges of the scar.

An example of this actually occurring in practice is shown in Fig. 8. A topographic image (with *x*,*y*,*z* coordinates) of a ball that has an abraded area is shown in Fig. 8(a). The abraded area is difficult to see at this magnification. Figure 8(b) is a difference image representing the difference between a sphere and the image in Fig. 8(a). The worn area now appears as a lump on the surface. Note that the unworn area of the difference image is not a flat plane, as would be expected. There is a sinewave-like pattern to it in the *x* direction. The *z* sensor of this particular profiling device is coupled through a clutch to a motor, which drives it in the *x* direction. The length in *x* of one period of the sinewave corresponds to the distance traveled by the *z* sensor during one revolution of the clutch. When a linear optical encoder was used to control data acquisition, the vertical size of this sinewave decreased by almost two orders of magnitude.

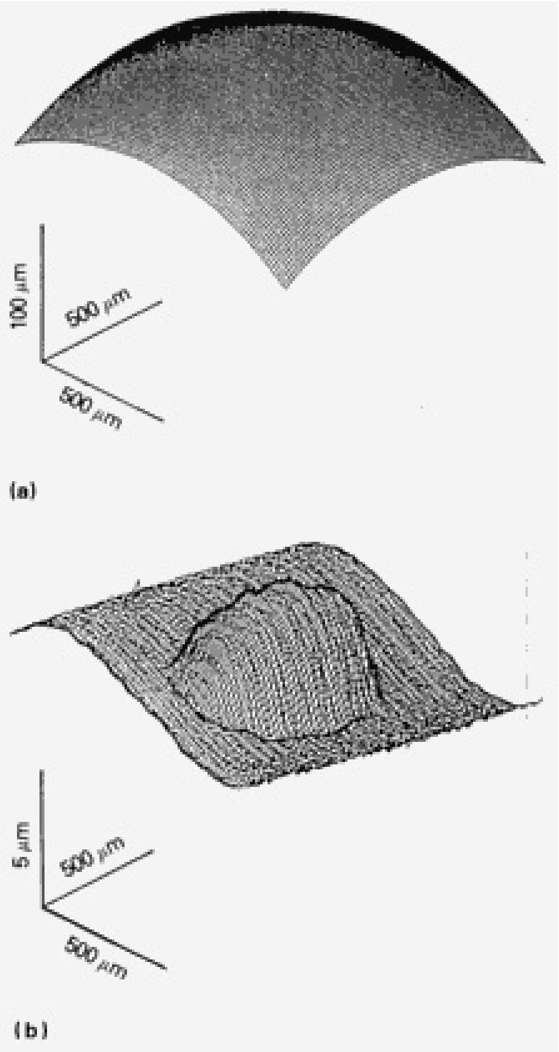


Fig. 8 (a) Topographic image with *x,y,z* coordinates of a ball with abraded area. (b) Difference image representing the difference between a sphere and the image in (a)

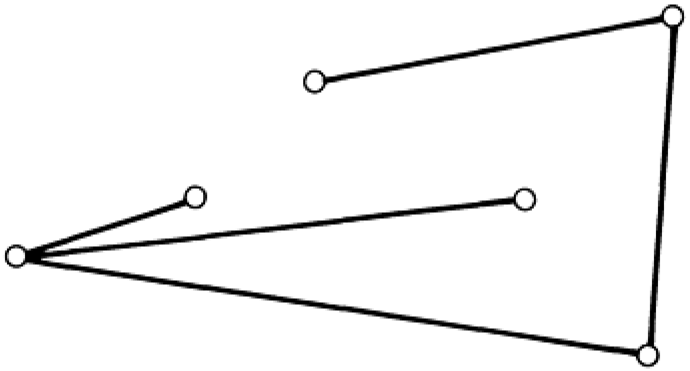
Curvature

There are a variety of techniques related to the determination of curvature for surfaces or for features. This information is useful when examining either slopes or the overall shape. It is generally desirable to use as many data points as possible in the determination of curvature to minimize the errors that are due to noise. The arrangement of neighboring pixels can be used for binary images (Ref 92). A spectral approach can be used (Ref 92, 93, 94, 95). Polynomials or circles can be fit to the data (Ref 15, 60, 96). The intersection of tangent lines is another technique (Ref 97). Curvature can be estimated from other computed parameters (Ref 92). Circular Hough transforms, discussed below, also provide a useful tool.

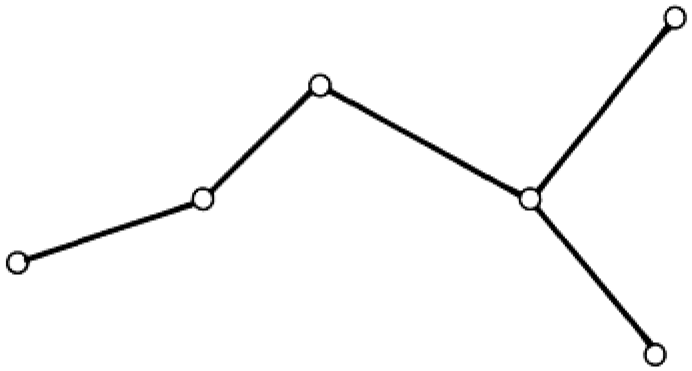
Hough Transforms and Pattern Matching. The Hough transform is one technique for locating shapes of known geometry in an image. For any shape, an appropriate function that maps an image or binary image onto a parameter space can be found. This mapping results in a sharp distribution of points in that parameter space around a coordinate representing the location of a selected reference point for that shape. Thus, the location of that shape has been located in the image. These mapping functions have been determined for lines (Ref 16, 98), circles (Ref 92, 98), and ellipses (Ref 98). There is also a method referred to as a general Hough transform. Used in computer vision systems, it can be applied to any arbitrary shape. Depending on how it is implemented, it can be either sensitive (Ref 98) or insensitive (Ref 99) to rotation. The sensitivity to noise and other error-producing effects have been studied in detail (Ref 100). This technique can be performed on either two-dimensional or three-dimensional data (Ref 101). A variety of other pattern-matching techniques also can be used (Ref 16, 102).

These techniques have the potential to enable an image analysis system to select individual features from surface images. These features could then be analyzed, manipulated, and classified individually. In one study, for example, an algorithm that learns which class a feature belongs to, according to the Fourier transform of its binary image, was developed (Ref 103). After a series of examples is given, the "typical" spectrum is automatically determined for each class of feature. The algorithm is then able to classify unknown features based on that learned "experience."

In another study, individual features were connected by a minimal spanning tree (Ref 33). A tree is a connected graph without closed loops, and a minimal spanning tree is a tree with the shortest possible total edge length. Figure 9 shows both a nonminimal and a minimal spanning tree, where each circle represents a feature. A histogram of the edge lengths was used to characterize the organization of the features on the surfaces studied. Trees are a very powerful tool and will be discussed again in the section below.



(a)



(b)

Fig. 9 (a) Nonminimal spanning tree. (b) Minimal spanning tree

Fractals, Trees, and Future Investigations

A fractal surface is one that contains a range of either regular or random geometric structures that exhibit some form of self-similarity over a range of scale (Ref 45). This self-similarity may be that the surface actually looks the same at a different magnification or that it produces the same statistics, such as roughness. A self-similar fractal (Fig. 10) is the "purest" fractal. It naturally appears self-similar, regardless of scale. At a magnification of 10×, a typical feature has a certain lateral and vertical size. If a section of this trace is selected and viewed at a higher magnification of 100×, then a typical feature has about the same lateral and vertical size as before. The process might be repeated at 1000×. Figure 10 is further discussed below.

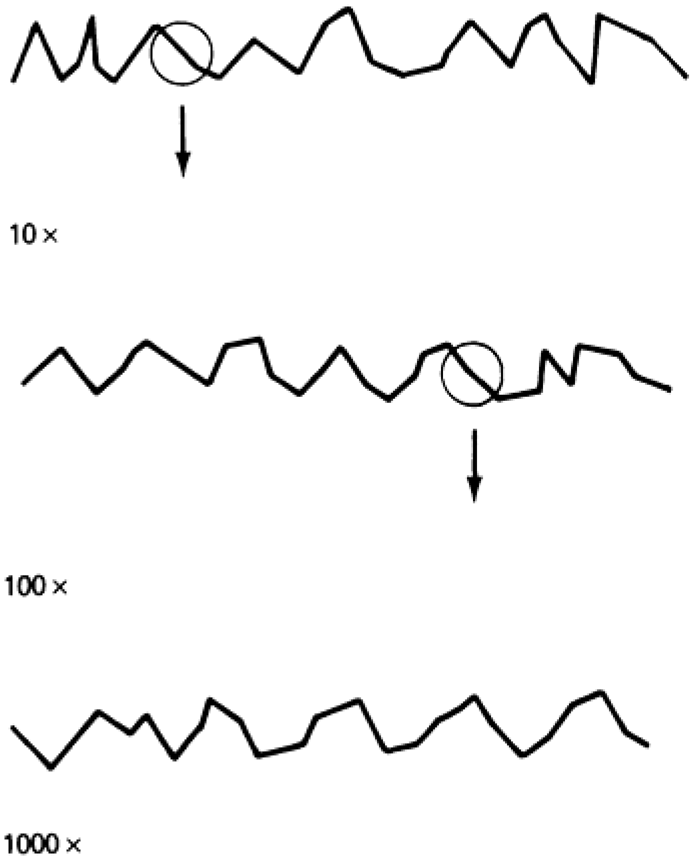


Fig. 10 Self-similar fractal

Self-similar fractals are described by their fractal dimension, which has a value from 1 to 2, for a single trace, and 2 to 3, for a surface. The integer part of the fractal dimension only indicates whether the data analyzed represent a trace (two-dimensional) or a surface (three-dimensional), and is not really important. The fractional part (on the right side of the decimal point) of the fractal dimension contains the important information.

In general, the higher the fractional part of the fractal dimension, the rougher the surface. However, many different methods of computing the fractal dimension have been derived, each yielding a different result (Ref 104). It has been shown, for example, that the fractal dimension of a fractured surface can have either a positive or a negative correlation with fracture toughness, depending on the details of how the fractal dimension is determined (Ref 105). Thus, care should be taken to know the details of how fractal dimensions are computed in an investigation. The range of sizes used in the calculation are very important and will be discussed later in this section of the article.

A self-affine fractal is only self-similar when expanded more in one direction than in another (Ref 106). Self-affine fractals require a second parameter, called the topothesy, which describes the scaling in one direction used to preserve self-similarity. Figure 11 shows an example. Like Fig. 10, sections of the image are selected and examined at progressively higher magnifications. In Fig. 10, the lateral and vertical size of a typical feature remained about constant for each magnification. However, in Fig. 11, the lateral size stays about constant while the vertical size increases. The vertical scale must therefore be compressed to maintain self-similarity.

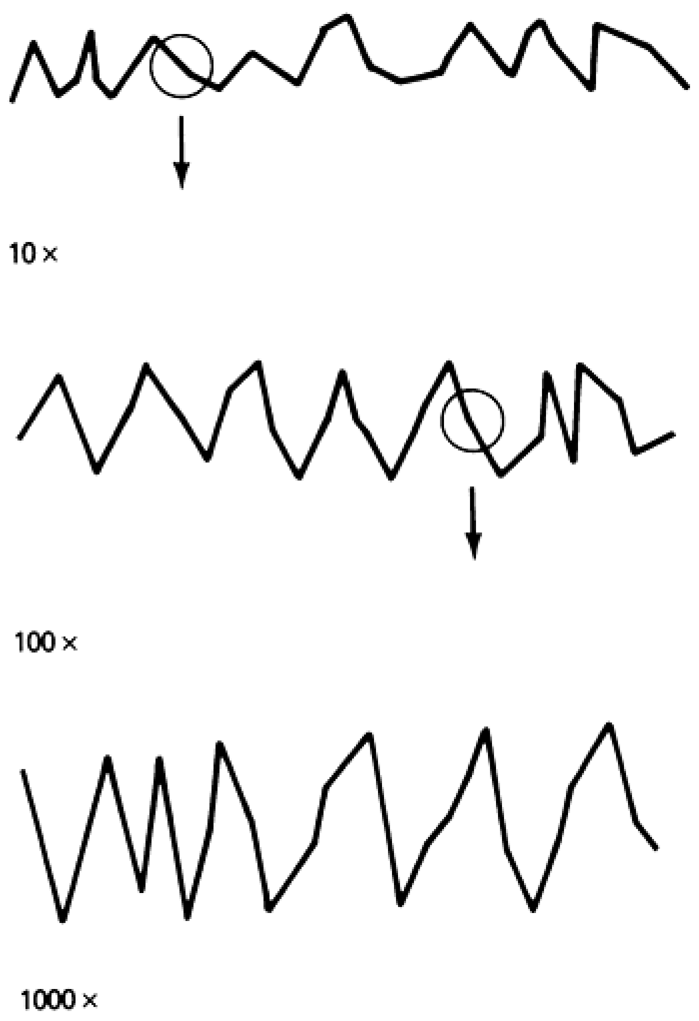


Fig. 11 Self-affine fractal

Unfortunately, self-affine fractals produce different values for most statistics at different magnifications. In fact, the variation of the standard deviation as a function of scale can be used to determine the topothesy (Ref 107). Single-valued functions can only be self-affine fractals, never self-similar. Because an image is a single-valuedfunction, images of fractal surfaces always appear as self-affine, even if the actual surface is self-similar. Thus, Fig. 10 could not actually occur unless the trace analyzed was not a single-valued function.

One example of this type of effect is a mountainous landscape on earth (Ref 107). When viewed from the top, contour lines of constant *z* height are often drawn in an *x,y* plane. These contour lines are not single-valued functions in *x* or *y* directions, and have been found to be self-similar. When *x,z* profiles of the same mountain are analyzed, they are single-valued functions of *x* in *z,* and are found to be self-affine fractals. There is also the possibility than an anisotropic surface may have a different topothesy, fractal dimension, or both, in different lateral directions (Ref 45). Some researchers have attempted to address this type of problem by using a matrix of fractal dimensions to describe surfaces (Ref 108).

Fractal behavior has been found in intensity images of surfaces of materials. The outlines of third-body wear particles in sliding (Ref 109, 110), martensite/austenite microstructures (Ref 111), and the growth of ion beam deposited alloy films (Ref 112) are examples. The topography of surfaces of materials has also been found to behave in a fractal manner, such as blasted steel panels (Ref 113, 114), coated surfaces (Ref 33), and fractured surfaces (Ref 105). One researcher examined worn rubber surfaces (Ref 110). The fractal dimensions of the surfaces were found to be limited to a finite size range and independent of the load, as long as the wear mechanism did not change. Surfaces of materials are always fractal only over some range of sizes. The largest scale possible is determined by the size of the specimen itself. The smallest scale is determined by the sizes of molecules. Any given profiling machine also covers only a certain range of sizes (Ref

22, 115, 116). If the size range of interest is too large for a single machine to characterize accurately, then images from several machines may be required.

Most surfaces appear to have several fractal dimensions, each over a different size range. One researcher describes a reasonably simple algorithm, which partially addresses this problem (Ref 49). It is assumed that there are two fractal dimensions in the image to be analyzed. The fractal dimension of the smaller, finer details is termed the textural fractal dimension. Another fractal dimension, which is found for the coarser, structural features, is termed the structural fractal dimension. If the two dimensions are not significantly different, then the surface is considered to have only one fractal dimension over the entire range of sizes analyzed. If they are different, then the scale of size where the surface changes from the one fractal dimension to the other is determined. Based on the results in Ref 107, a similar approach can be performed for topothesy.

It is possible for a surface to have different fractal dimensions and/or topothesy in different areas occurring simultaneously, even within the same size range. A groove on a worn surface may have different characteristics than a lump, for example. Though not yet documented for topographic images of materials surfaces, it is conceivable that such a phenomenon can occur. Although it is difficult to thoroughly verify this observation with current techniques, the exploration of such phenomena will now be discussed, because they serve as good examples of how future investigations might be performed.

Consider a severely worn surface that has large grooves, ridges, holes, and lumps. Each of these types of features can have smaller grooves, ridges, holes, and lumps. It may be that the grooves of various sizes, when considered separately from the other types of features, have one fractal dimension, whereas lumps have another. This could theoretically be tested by generating four new images from the original image. One image would consist of only the grooves, one of only the ridges, one of only holes, and one of only lumps. This might be accomplished using a multiscale pattern-matching algorithm (Ref 117). The fractal dimension of these images could then be compared and any differences characterized.

It is also of interest to determine if a large groove and a large lump have the same "mix" of smaller features on them. There are several ways to investigate this. One is to generate a new image, where each pixel represents the fractal dimension of the original image immediately around that location. How the fractal dimension changes as a function of lateral position can then be studied. Such a procedure has been used in medical imaging (Ref 118) and for studying the sea floor (Ref 106). The same procedure might be applied to other parameters, as well.

There is, of course, the issue of how large a sampling area each pixel in the new image should represent. A technique termed "adaptive mask selection" attempts to determine the optimal sampling area for each pixel (Ref 119).

A second approach for studying the types of smaller features that are contained on larger ones uses a multistep process. First, select a typical large groove and generate a second image of just that groove and all its internal structure. Next, filter out the longer wavelengths of the large groove to generate a third image of only the smaller internal features. These steps can then be repeated for large ridges, holes, and lumps. The topographies of the smaller features within the larger features could then be compared.

Additionally, this process could then be repeated recursively on the smaller features to determine what each of them contains. This would result in a tree structure, known as a relationship tree (Ref 120). Figure 12 depicts how part of such a tree might look. Note that the overall size of each feature is also recorded in the tree. As noted previously, much of computer science is devoted to the manipulation and classification of trees, making this form of representation very powerful.

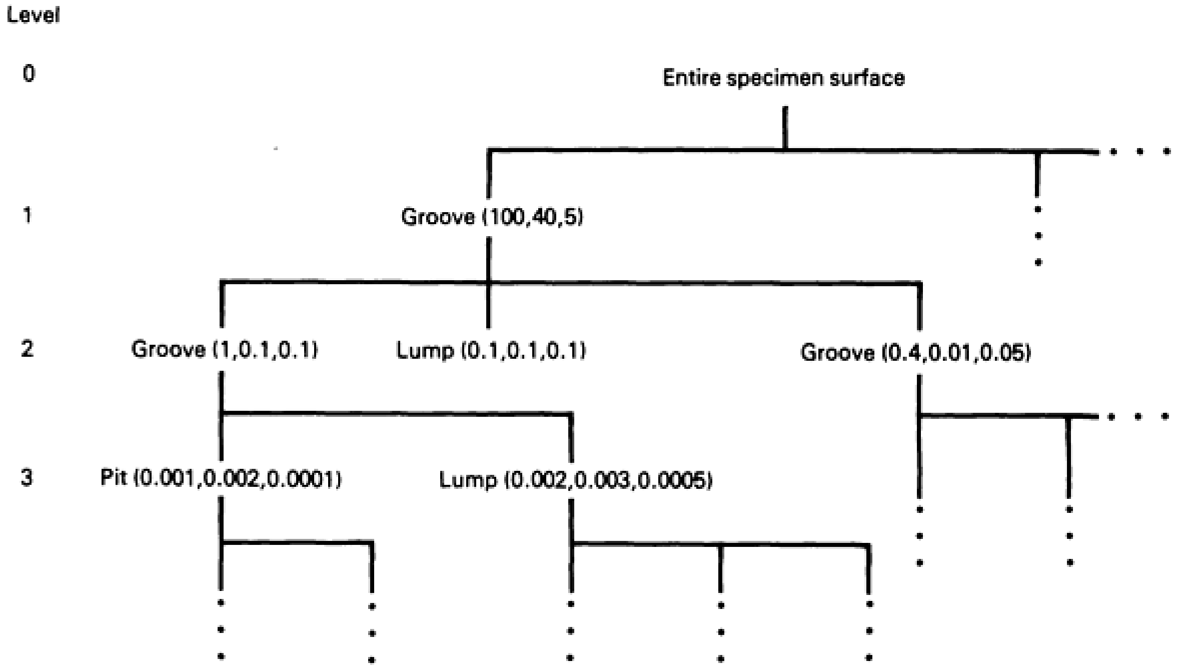


Fig. 12 Example of portion of relationship tree of larger features, each containing smaller features. Each node contains information denoting both type of feature and overall feature size in micrometers.

References

1. T.R. Thomas, Ed., *Rough Surfaces,* Longman, 1982
2. A. Ya Grigor'ev, N.K. Myshkin, O.V. Kholodilov, Surface Microgeometry Analysis Methods, *Sov. J. Frict. Wear,* Vol 10 (No. 1), 1989, p 138-155
3. J.C. Russ, *Computer-Assisted Microscopy, The Measurement and Analysis of Images,* Plenum Press, 1990
4. *Image Analysis and Metallography,* Vol 17, *Microstructural Science,* ASM International, 1989
5. B.H. Kaye, *Direct Characterization of Fine Particles,* John Wiley & Sons, 1981
6. P.C. Montgomery, Nanoscopy: Nanometer Defect and Analysis by Computer Aided 3D Optical Imaging, *Nanotechnology,* Vol 1, 1990, p 54-62
7. A.N. Tabenkin and F.G. Parsons, Surface Analysis for Measurement of Pigment Agglomeration in Rubber, *Metrology and Properties of Engineering Surfaces, 4th International Conference,* 1988, p 301-308
8. C.F. Fahl, Motif Combination--A New Approach to Surface Profile Analysis, *Wear,* Vol 83, 1982, p 165-179
9. "Results of the Work Carried Out by Member Body of France to Perform the Resolution 14 of the ISO/TC57/SC4 in Bad Durrheim, FRG (June 1988)," presented at ISO meeting at National Institute of Standards and Technology, 1990
10. M. Chuard, D. Rondot, and J. Mignot, Numerical Filtering used in Microtopographical Surface Measurement, *Wear,* Vol 97, 1984, p 267-274
11. U. Scheider, A. Steckroth, N. Rau, and G. Hubner, An Approach to the Evaluation of Surface Profiles by Separating Them into Functionally Different Parts, *Metrology and Properties of Engineering Surfaces, 4th International Conference,* 1988, p 343-356
12. E.C. Teague, F.E. Scire, and T.V. Vorburger, Sinusoidal Profile Precision Roughness Specimens, *Wear,* Vol 83, 1982 p 61-73
13. B.J.H. Verwer, F.C.A. Groen, R.J. Ekkers, and R. Ellens, Image Processing on Personal Computers, *IEEE*

International Conference on Systems Engineering, 1989, p 183-186

14. A. Johnson, Scientific Image Processing on the Macintosh II, *IEEE International Conference on Systems Engineering,* 1989, p 179-182
15. T. Pavlidis, Algorithms for Graphics and Image Processing, *Computer Science Press,* 1982
16. M. James, *Pattern Recognition,* John Wiley & Sons, 1988
17. F. Tomita and S. Tsuji, *Computer Analysis of Visual Textures,* Kluwer Academic Publishers, 1990
18. M.B. Weppner and M. Young, "Image Processing for Optical Engineering Applications," Report NBSIR 87-3065, National Institute of Standards and Technology, 1987
19. T. Vaughan, *Using HyperCard, From Home to HyperTalk,* Que Corporation, 1988
20. N.-S. Chang, *Image Analysis and Image Database Management,* UMI Research Press, 1981
21. J.A. Ogilvy and J.R. Foster, Rough Surfaces: Gaussian or Exponential Statistics, *J. Appl. Phys. D: Applied Physics,* Vol 22, 1989, p 1243-1251
22. B. Marchon, S. Vierk, N. Heiman, R. Fisher, and M.R. Khan, "Significance of Surface Roughness Measurements. Application to the Tribology of the Head/Disc Interface," Special Publication 26, Society of Tribologists and Lubrication Engineers, 1989, p 71-80
23. E.C. Teague, On Measuring the Root-Mean-Square Value of a Finite Record Length Periodic Waveform, *J. Res. NIST,* Vol 94, 1989, p 367-371
24. D.G. Chetwynd, The Digitization of Surface Profiles, *Wear,* Vol 57, 1979, p 137-145
25. I. Sherrington and E.H. Smith, Fourier Models of the Surface Topography of Engineered Components, *Metrology and Properties of Engineering Surfaces, 4th International Conference,* 1988, p 167-182
26. M. Razinger and M. Novic, Reduction of the Information Space for Data Collections, *PCs for Chemists,* 1990, p 89-103
27. T. Tsukada and K. Sasajima, An Optimum Sampling Interval for Digitizing Surface Asperity Profiles, *Wear,* Vol 83, 1982, p 119-128
28. B.E. Klamecki, Information Content of Surfaces Profiles, Surface Profile Measurements, and Profile Characterizations, *Wear,* Vol 130, 1989 p 289-300
29. E.H. Smith and W.M. Walmsley, Walsh Functions and Their Use in the Assessment of Surface Texture, *Wear,* Vol 57, 1979, p 157-166
30. M.I. Yolles, E.H. Smith, and W.M. Walmsley, Walsh Theory and Spectral Analysis of Engineering Surfaces, *Wear,* Vol 83, 1982, p 151-164
31. P. Spilling, Ed., *Numerical Techniques,* Vol 5, *Characterization of High-Temperature Materials,* The Institute of Metals, 1989
32. J.-H. Lee and J.C. Russ, Metrology of Microelectronic Devices by Stereo SEM, *J. Comput.-Assist. Microsc.,* Vol 1 (No. 1), 1989, p 79-90
33. M. Rasigni, G. Rasigni, F. Varnier, C. Dussert, and A. Liebaria, Statistical Analysis of Random and Pseudo Random Rough Surfaces, *Surface Measurement and Characterization, SPIE Proceedings Series,* Vol 1009, 1988, p 68-76
34. I.-M. Feng, A New Approach in Interpreting the Four-Ball Wear Results, *Wear,* Vol 5, 1962, p 275-288
35. E.P. Whitenton and P.J. Blau, A Comparison of Methods for Determining Wear Volumes and Surface Parameters of Spherically Tipped Sliders, *Wear,* Vol 124, 1988, p 291-309
36. N. Ahuja and B.J. Schachter, Pattern Models, John Wiley & Sons, 1983
37. B. Snaith, S.D. Probert, and R. Pearce, Characterization of Laser-Textured Cold-Rolled Steel Sheets, *Wear,* Vol 109, 1986, p 87-97
38. T. Tsukada and T. Kanada, Evaluation of Two- and Three-Dimensional Surface Roughness Profiles and their Confidence, *Wear,* Vol 109, 1986, p 69-78
39. K.J. Stout and J. Davis, The Specification of Surface Finish Tolerance for the Control of Manufacture of Engineering Surfaces, *Wear,* Vol 109, 1986, p 181-193
40. E.P. Whitenton and D.E. Deckman, Measuring Matching Wear Scars on Balls and Flats, *Surf. Topogr.,*

Vol 2 (No. 3), 1989, p 205-222

41. F. George and S.J. Radcliffe, Automated Wear Measurement on a Computerized Profilometer, *Wear,* Vol 83, 1982, p 327-337
42. E.J. Davis, K.J. Stout, and P.J. Sullivan, The Scope of Three Dimensional Topography, *Ind. Metrol.,* Vol 1, 1990, p 193-216
43. E.P. Whitendon, M.B. Peterson, and L.K. Ives, Method for Quantitative Measurement of Galling Damage, *Metal Transfer and Galling in Metallic Systems,* The Metallurgical Society, 1987, p 155-170
44. G.R. Dickson, R. McIlhagger, and P.P. Miller, The Effect of Mould Surface Finish and Processing Conditions on the Surface Characteristics of Polymeric Injection Mouldings, *Metrology and Properties of Engineering Surfaces, 4th International Conference,* 1988, p 395-401
45. T.R. Thomas and A.P. Thomas, Fractals and Engineering Surface Roughness, *Metrology and Properties of Engineering Surfaces, 4th International Conference,* 1988, p 1-10
46. W. Watson and A. Woods, The Three Dimensional Representation of Engineering Surfaces, *Metrology and Properties of Engineering Surfaces, 4th International Conference,* 1988, p 11-28
47. V.M. Izraelev, *2D Versus 3D Surface Roughness Measurement,* Special Publication 26, Society of Tribologists and Lubrication Engineers, p 28-32
48. A. Bengtsson and A. Ronnberg, The Absolute Measurement of Running-In, *Wear,* Vol 109, 1986, p 329-342
49. R. Crutzburg and E. Ivanov, Fast Algorithm for Computing Fractal Dimensions of Image Segments, *Recent Issues in Pattern Analysis and Recognition,* Springer-Verlag, 1989, p 42-51
50. C. Roques-Carmes, D. Wehbi, J.F. Quiniou, and C. Tricot, Modeling Engineering Surfaces and Evaluating their Non-Integer Dimension for Application in Material Science, *Metrology and Properties of Engineering Surfaces, 4th International Conference,* 1988, p 237-247
51. S.R. Lange and B. Bushan, Use of Two- and Three-Dimensional, Noncontact Surface Profiler for Tribology Applications, *Metrology and Properties of Engineering Surfaces, 4th International Conference,* 1988, p 205-218
52. Y. Tanimura, E.C. Teague, F.E. Scire, R.D. Young, and T.V. Vorburger, Graphical Signatures for Manufactured Surfaces, *J. Lubr. Technol.,* Vol 104, 1982, p 533-537
53. R.S. Sayles and T.R. Thomas, Surface Topography as a Nonstationary Random Process, *Nature,* Vol 271 (No. 5644), 1978, 431-434
54. H.S. Nielsen, New Approaches to Surface Roughness Evaluation of Special Surfaces, *Precis. Eng.,* Vol 10 (No. 4), 1988, p 209-213
55. B. Nowicki, Multiparameter Representation of Surface Roughness, *Wear,* Vol 102, 1985, p 161-176
56. T.G. King and T.A. Spedding, On the Relationship Between Surface Profile Height Parameters, *Wear,* Vol 83, 1982, p 91-108
57. A. Bruzzone, P.M. Lonardo, and G. Vernazza, Cluster Analysis and Representation for Topology of Mechanically Worked Surfaces, *Image Analysis and Image Processing,* Plenum Press (Rapalto, Italy), 1985
58. P.J. Scott, Developments in Surface Texture Measurement, *Metrology and Properties of Engineering Surfaces, 4th International Conference,* 1988, p 373-384
59. W. Prater, Testing Ball Bearings Can Prevent Noise In Head/Disk Assemblies, Walter Prater, *Computer Technology Review,* Winter 1984, p 65-70
60. T.S.R. Murthy, B. Raghunatha Rao, and S.Z. Abdin, Evaluation of Spherical Surfaces, *Wear,* Vol 57, 1979, p 167-184
61. A.L. Gauler, The Characterization of Spherical Surfaces, *Wear,* Vol 83, 1982, p 109-118
62. V. Soundararajan and V. Radhakrishnan, An Investigation on the Roundness and Side-wall Errors in Ultrasonic Drilling, *Surf. Topogr.,* Vol 2 (No. 3), 1989, p 233-246
63. P. Newman, S.J. Radcliffe, and J. Skinner, The Accuracy of Profilometric Wear Volume Measurement on the Rough LCIB-Coated Surfaces of an Articulating Joint, *Surf. Topogr.,* Vol 2, 1989, p 59-77

64. P.I. Lacey, S.M. Hsu, *et al.,* "Wear Mechanisms of Valves and Valve Seat Inserts in a Gas-Fired Reciprocating Engine," NISTIR 90-4264, National Institute of Standards and Technology

65. J. Peklenik and M. Kubo, A Basic Study of a Three-Dimensional Assessment of the Surface Generated in a Manufacturing Process, *Annals of the CIRP,* Vol XVI, 1968, p 257-265

66. M. Kubo and J. Peklenik, An Analysis of Micro-Geometrical Isotropy for Random Surface Structures, *Annals of the CIRP,* Vol XVI, 1968, p 235-242

67. T.S.R. Murthy and G.C. Reddy, Different Functions and Computations for Surface Topography, *Wear,* Vol 83, 1982, p 203-214

68. *Surface Texture (Surface Roughness, Waviness, and Lay),* The American Society of Mechanical Engineers, ANSI/ASME B46.1, 1985

69. D.J. Whitehouse and R.E. Reason, *The Equation of the Mean Line of Surface Texture found by an Electronic Wave Filter,* Rank Taylor Hobson, 1965

70. T.V. Vorburger and J. Raja, *Surface Finish Metrology Tutorial,* NISTIR 89-4088, National Institute of Standards and Technology, 1989

71. P.J. Sullivan and N. Luo, The Use of Digital Techniques for Error Correction in Surface Roughness Measurement, *Surf. Topogr.,* Vol 2, 1989, p 143-157

72. F.X. Wang, P. Lacey, S.M. Hsu, and R.S. Gates, "A Study of the Relative Surface Conformity Between Two Surfaces in Sliding Contact," 90-Trib-61, ASME/STLE

73. J. Bielle, Functional Needs, Machining Conditions, and Economics of Surface Finish, *Precis. Eng.,* 1985, p 31-38

74. B. Bhushan and M.F. Doener, "Role of Mechanical Properties and Surface Texture in the Real Area of Contact of Magnetic Rigid Disks," 88-Trib-39, The American Society of Mechanical Engineers

75. T. Merriman and J. Kannel, "Analyses of the Role of Surface Roughness on Contact Stresses Between Elastic Cylinders With and Without Soft Surface Coatings," 88-Trib-54, The American Society of Mechanical Engineers

76. M.N. Webster, E. Ioannides, and R.S. Sayles, The Effect of Topographical Defects on the Contact Stress and Fatigue Life in Rolling Element Bearings, *Mechanisms and Surface Distress,* Butterworths, 1985, p 207-221

77. A. Ishibashi, S. Tanaka, and S. Ezoe, Changes in Contact Surfaces and Surface Durability of Highly Loaded Gears with Roughnesses of 0.1 to 15 μm R_{max}, *Mechanisms and Surface Distress,* Butterworths, 1985, p 339-346

78. M.N. Webster and R.S. Sayles, A Numerical Model for the Elastic Frictionless Contact of Real Rough Surfaces, *J. Tribol.,* Vol 108 (No. 3), 1986, p 314-320

79. M.A. West and R.S. Sayles, "A 3-Dimensional Method of Studying 3-Body Contact and Stress on Real Rough Surfaces," 14th Leeds Lyon Symposium, 1987

80. M.N. Webster, M.A. West, and R.S. Sayles, A. Method of Three-Dimensional Topography Measurement and Analysis on Arcuate Surfaces, *Wear,* Vol 109, 1986, p 385-399

81. M.A. West, M.N. Webster, and R.S. Sayles, "A New Method for Rough Surface Contact Analysis," C161/87, IMechE, 1987

82. J.I. McCool, Predicting Flash Temperature at Microcontacts, *Metrology and Properties of Engineering Surfaces, 4th International Conference,* 1988, p 267-280

83. V.S. Poroshin, On the Three-Dimensional Contact Problem with Quadratic Dependence of the Rough Layer Displacements on Pressure, *Sov. J. Frict. Wear,* Vol 10 (No. 1), 1989, p 66-68

84. A. Williams and N. Idrus, Detection and Measurement of Damage to Surfaces After Static Loading, *Wear,* Vol 57, 1979, p 281-291

85. J. Kagami, T. Hatazawa, K. Yamada, and T. Kawaguchi, Three-Dimensional Observation and Measurement of Worn Surfaces, *Surf. Topogr.,* Vol 1, 1988, p 47-60

86. R. DeLong and M.C. Bramwell, Matching an Initial Surface to a Worn Surface Subject to Rotation and Translation, *The Mathematics of Surfaces II,* Clarendon Press, 1987

87. J.L. Lauer and S.S. Fung, Microscopic Contour Changes of Tribological Surfaces by Chemical and Mechanical Action, *ASLE Trans.,* Vol 26 (No. 4), 1982, p 430-436

88. J.L. Lauer and S.S. Fung, Topological Reaction Rate Measurements Related to Scuffing, *ASLE Trans.,* Vol 27 (No. 4), 1983, p 228-294

89. T.C. Buttery and M.S. Hamed, A Study of Geometric and True Leveling in Electroplating Using Stylus Methods, *Wear,* Vol 57, 1979, p 207-215

90. A.S. Morris, *Principles of Measurement and Instrumentation,* Prentice-Hall, 1988, p 98-101

91. B.A. Bell, Ed., *Digital Methods in Waveform Metrology,* NBS Special Publication 707, National Institute of Standards and Technology, 1985

92. J.C. Russ, Automatic Methods for the Measurement of Curvature of Lines, Features, and Feature Alignment in Images, *J. Comput.-Assist. Microsc.,* Vol 1 (No. 1), 1989, p 39-77

93. H. Moalic, J.A. Fitzpatrick, and A.A. Torrance, "A Spectral Approach to the Analysis of Rough Surfaces," 88-Trib-57, The American Society of Mechanical Engineers

94. H. Moalic, J.A. Fitzpatrick, and A.A. Torrance, The Correlation of the Characteristics of Rough Surfaces with their Friction Coefficients, *Proc. Inst. Mech. Eng.,* Vol 201 (No. C5), 1987, p 321-329

95. M.N.H. Damir, Approximate Harmonic Models for Roundness Profiles, *Wear,* Vol 57, 1979, p 217-225

96. R.A. Liming, *Mathematics for Computer Graphics,* Aero Publishers, Inc., 1979

97. A. Milano, F. Perotti, S.B. Serpico, and G. Veranazza, Detection of Arcs in Workpiece Images, *Recent Issues in Pattern Analysis and Recognition,* Springer-Verlag, 1989, p 324-337

98. D.H. Ballard and C.M. Brown, *Computer Vision,* Prentice-Hall, Inc., 1982

99. A. Arbuschi *et al.,* Relocation and Location of Mechanical Parts using the Hough Technique, *Digital Image Analysis,* 1984, p 373-379

100. W.E.L. Grimson and D.P. Huttenlocher, On the Sensitivity of the Hough Transform for Object Recognition, IEEE *Transactions on Pattern Analysis and Machine Intelligence,* Vol 12 (No. 3), 1990, p 255-274

101. W.-H. Lou and A.P. Reeves, Three-Dimensional Generalized Hough Transform for Object Identification, *Intelligent Robots and Computer Vision VIII: Algorithms and Techniques,* Society of Photo-Optical Instrumentation Engineers, 1990, p 363-374

102. D.B. Goldgof, T.S. Huang, and H. Lee, Feature Extraction and Terrain Matching, *Proceedings CVPR '88: The Computer Society Conference on Computer Vision and Pattern Recognition,* 1988, p 899-904

103. G. Bonifazi, P. Massacci, and G. Patrizi, Alternative Feature Selection Procedures for Particle Classification by Pattern Recognition Techniques, *Recent Issues in Pattern Analysis and Recognition,* Springer-Verlag, 1989, p 365-376

104. H. Takayasu, Notes on Fractal Dimension, *Fractals in the Physical Sciences,* Manchester University Press, 1990, p 141-163

105. C.W. Lung and S.Z. Zhang, Fractal Dimension of the Fractured Surface of Materials, *Phys. D,* Vol 38, 1989, p 242-245

106. A. Malinverno, Segmentation of Topographic Profiles of the Seafloor Based on a Self-Affine Model, *IEEE J. Ocean. Eng.,* Vol 14 (No. 4), 1989, p 348-359

107. M. Matsushita and S. Ouchi, On the Self-Affinity of Various Curves, *Phys. D,* Vol 38, 1989, p 246-251

108. H. Kaneko, "A Generalized Fractal Dimension and its Application to Texture Analysis--Fractal Matrix Model," CH2673-2/89/0000-1711, IEEE, 1989, p 1711-1714

109. D. Wehbi, C. Roques-Carmes, A. Le Mehaute, and C. Tricot, *Fractal Aspects of Materials: Disordered Systems: Extended Abstracts,* Materials Research Society, 1987, p 185-187

110. P.R. Stupak and J.A. Donovan, Fractal Analysis of Rubber Wear Surfaces and Debris, *Fractal Aspects of Materials: Disordered Systems: Extended Abstracts,* Materials Research Society, 1987, p 199

111. E. Hornbogen, The Fractal Nature of Martensite/Austenite Microstructures, *Martensitic Transformations Ptl,* 1989, p 131-138

112. L.J. Huang, B.X. Liu, J.R. Ding, and H.D. Li, Formation of Fractal Patterns in Nickel Based Alloy Films

by Ion Beam Method, *Fractal Aspects of Materials: Disordered Systems: Extended Abstracts,* Materials Research Society, 1987, p 188-190

113. J.W. Martin and D.P. Bentz, Fractal-Based Description of the Roughness of Blasted Steel Panels, *J. Coat. Technol.,* Vol 59 (No. 745), 1987, p 35-41
114. J.W. Martin and D.P. Bentz, Roughness Measures of Blasted Steel Surfaces Remotely Imaged with a Thermographic Camera, *Corrosion Protection by Organic Coatings,* Proceedings Vol 87-2, The Electrochemical Society, Inc., 1987, p 179-196
115. F.F. Ling, "The Possible Role of Fractal Geometry in Tribology," Reprint 88-TC-2A-1, Society of Tribologists and Lubrication Engineers
116. E.C. Teague, F.E. Scire, S.M. Baker, and S.W. Jensen, *Wear,* Vol 83, 1982, p 1-12
117. A. Taza and C.Y. Suen, Discrimination of Planar Shapes Using Shape Matrices, *IEEE Transactions on Systems, Man, and Cybernetics,* Vol 19 (No. 5), 1989, p 1281-1289
118. C.-C. Chen, J.S. Daponte, and M.D. Fox, Fractal Feature Analysis and Classification in Medical Imaging, *IEEE Transactions on Medical Imaging,* Vol 8 (No. 2), 1989, p 133-142
119. F. Arduini, C. Dambra, S. Dellepiane, S.B. Serpico, G. Vernazza, and R. Viviani, "Fractal Dimension Estimation by Adaptive Mask Selection," CH2561-9/88/0000-1116, IEEE, 1988
120. G.R. Wilson and B.G. Batchelor, Algorithm for Forming Relationships Between Objects in a Scene, *IEEE Proceedings,* Vol 137, Pt E, No. 2, 1990, p 151-153

Confocal Microscopy

Monica A. Schmidt, Martin Marietta Energy Systems, Inc.; Robert D. Compton, NORAN Instruments, Inc.

Introduction

CONFOCAL MICROSCOPY is a relatively new technique for surface topography imaging that has applications in characterizing worn and damaged surfaces. Confocal imaging uses a special optical microscope, sometimes called a confocal scanning optical microscope (CSOM), which is interfaced to an image analyzer to allow quantitative measurements of surface topography, in both lateral and vertical dimensions. Optical images that maintain clear focus over a large depth of field can be produced. Surface roughness measurements can be made on opaque samples, and subsurface structure can be clearly imaged and characterized in optically translucent samples.

Image analysis techniques allow three-dimensional reconstruction and sectioning of surfaces (Ref 1), and generation of stereo images. Confocal microscopy is ideal for the imaging and analysis of irregular structures, such as fracture surfaces. Surfaces with large differences in height are difficult to interpret using conventional light microscopy, but are ideal for confocal microscopy.

The term confocal simply means "single focus." In a confocal microscope, confocality is achieved through the use of pinhole optics that prevent out-of-focus light from reaching the image plane. The sample is illuminated through an objective lens with a pinpoint of light, and a pinhole aperture is placed in the reflected light path. Light reflected from the sample at the focal plane of the objective lens passes back through the lens, through the pinhole, and forms an image of the illuminated spot. Reflected light from other regions of the sample is blocked by the aperture. An image of the sample is created by moving either the sample or the light source in an appropriate scan pattern, and either recording or viewing the resulting signals.

Confocal images, also called optical sections, have very good resolution and sharp contrast levels because only light reflected at the focal plane of the objective lens is imaged. Light reflected from sample features distant from the focal plane of the objective lens is blocked from the image, so that defocused regions remain dark and cannot deteriorate the resolution of the image.

Confocal microscopy offers new capabilities that can be applied to wear and abrasion studies. Surface topography can be characterized visually and quantitatively using microscopy and image analysis techniques. Scratches, gouges, and other forms of surface damage can be measured without physically contacting the surface, eliminating the risk of creating additional damage during the measurement. Induced subsurface damage in translucent materials can also be analyzed, which provides new opportunities for materials evaluation in failure analysis or from controlled wear experiments.

Confocal microscopes have found broad use in semiconductor (Ref 2), biological, and other materials applications (Ref 3). In contrast to electron microscopy, samples do not have to be electrically conductive or vacuum compatible. Applications that utilize the three-dimensional capabilities of confocal microscopy include studying internal features of polymers and inspecting integrated circuits. Microscopy of biological materials has benefited dramatically from the application of this new tool, particularly the determination of fluorescence-stained cellular and tissue structures. In these applications, confocal imaging can show the internal structures of translucent specimens that have not been dissected. Real-time imaging makes it possible to study living tissue.

Development of Confocal Microscopy. The first patent on confocal microscopy was filed by Minsky in 1957 (Ref 4). In this patent, a pinpoint of light was focused onto a specimen, reflected light was focused onto a pinhole, and light passed by the pinhole was optically coupled to the image plane. An image was created by electromechanically moving the sample in a scan pattern and using the illumination passing through the pinhole to produce an image on a long-persistence cathode-ray tube.

The next major step was taken in the late 1960s by Petran and Hadravsky (Ref 5), whose designed centered around the use of a Nipkow disk (Ref 6) (Fig. 1). The apertures of this disk simultaneously formed multiple points of light and served as pinholes to produce confocal imaging. By spinning the disk rapidly, this tandem scanning microscope (TSM) produced a real-time confocal image.

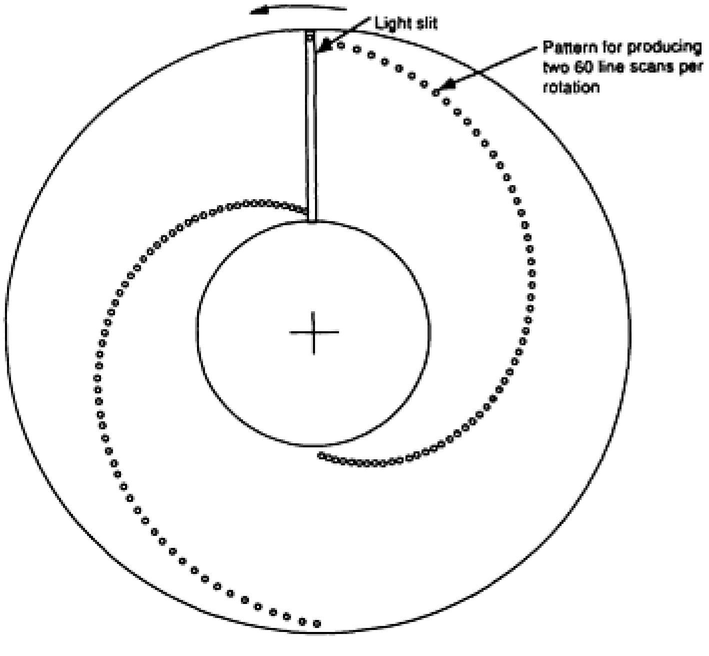


Fig. 1 Classic Nipkow disk, where light passing through the slit and the disk apertures is detected to produce a line scan of the image. If hundreds of aperture tracks are used to form a full symmetrical pattern, real-time imaging can be achieved.

In the 1970s, lasers were introduced as point sources. In 1987, the modern laser confocal scanning microscope (LCSM) was developed by Aslund *et al.* (Ref 7). This microscope employed high-speed galvanometer scanning and produced images in 1 to 2 seconds. High-performance commercial systems of both the TSM and the LCSM variety appeared in the marketplace shortly afterward. The most recent development, a real-time LCSM that was commercially introduced in 1990, features acousto-optic deflection of the laser beam.

Acknowledgements

This research was sponsored by the U.S. Department of Energy, Assistant Secretary for Conservation and Renewable Energy, Office of Transportation Technologies, as part of the High-Temperature Materials Laboratory User Program, under contract DE-AC05-84OR21400 with Martin Marietta Energy Systems, Inc.

The authors would like to acknowledge the support of Dr. Peter J. Blau and Dr. Charles S. Yust, who provided the specimens to demonstrate applications for confocal microscopy and who volunteered their time to review this article.

Principles of Confocal Microscopy

A confocal microscope is a specialized form of a reflected-light microscope that can measure both vertical and lateral dimensions of a specimen. Its image, called an optical section, displays a planar view of the specimen, centered about the focal plane of the objective lens and oriented perpendicular to the optical axis (Fig. 2).

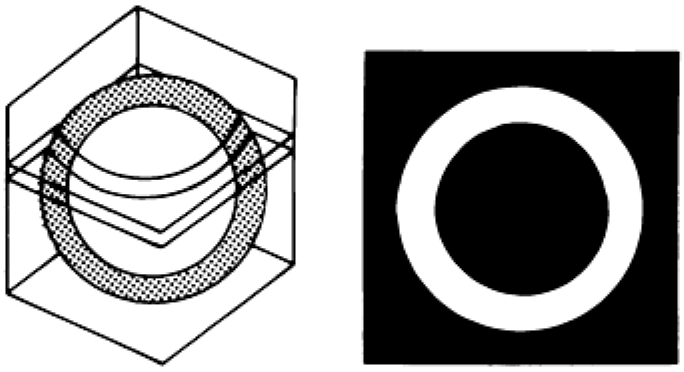


Fig. 2 Optical slice of hollow sphere

The optical section has both high resolution and contrast, because the confocal design of the microscope blocks light reflected from specimen regions distant from the focal plane of the objective lens. Thus, regions that are in focus (near the focal plane) appear bright, but regions that are out of focus (distant from the focal plane) appear dark in this optical section. Exceptionally clear images of the surface features of opaque samples can be obtained, as can internal features of transparent samples, because defocused light is removed from the image.

In a basic confocal microscope, confocality is achieved by collimating the light source, focusing light through the objective lens to form a spot on the focal plane, and focusing reflected light through the same objective lens onto an aperture (Fig. 3). Light that passes through the aperture is focused onto a detector, where it forms one point of an image. Most of the light from other areas, including scattered, reflected, and fluorescent light from out-of-focus planes, is blocked by the aperture and cannot degrade the image. Thus, the information contained in the imaged spot is limited to a narrow elevation range centered around the focal plane of the objective. Lateral scanning of either the sample or the light source is necessary to build up a complete optical section.

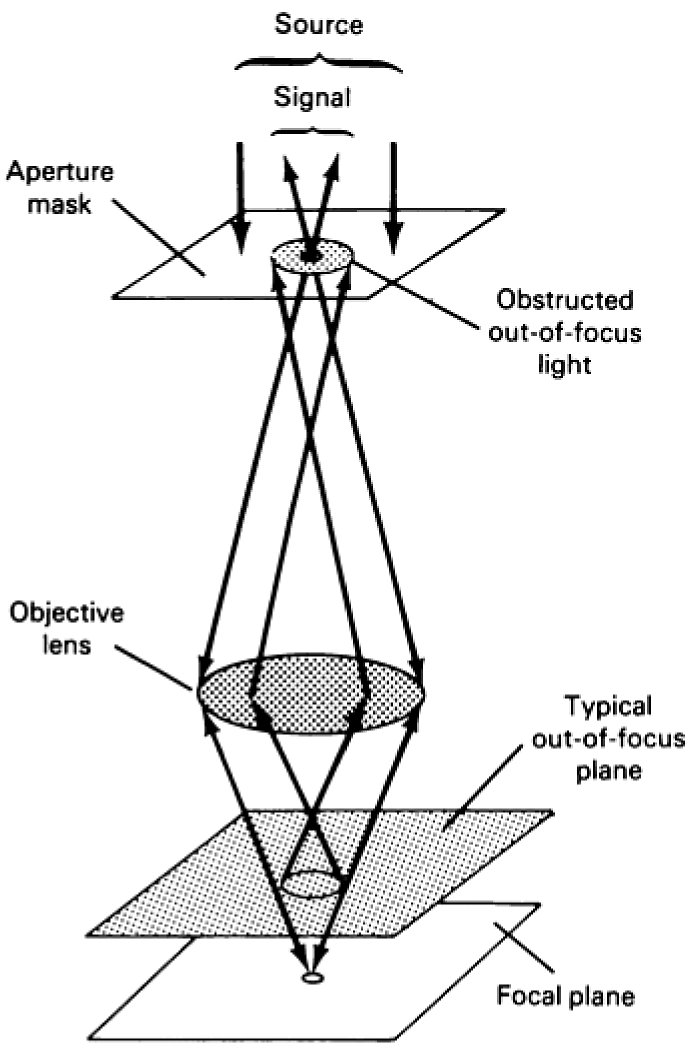


Fig. 3 Confocal principal

The vertical resolution of the confocal optical section is the function of several factors, including aperture size, wavelength of the light source, and numerical aperture of the objective lens. Thus, the operator has some degree of control, primarily by his selection of the objective lens. The vertical resolution is proportional to the square of the numerical aperture (Ref 8), and reaches a limit of about 0.4 μm (16 μin.) for a dry objective lens with a numerical aperture of 0.95. The transverse resolution also is controlled by selection of the objective lens, but is proportional to the first power of the numerical aperture. Thus, the availability of suitable objective lenses can limit applicability of the microscope, depending on the size of features to be analyzed.

The intensity of reflected light in each optical section varies inversely with the distance of the sample feature from the focal plane of the objective lens. This characteristic allows image analysis techniques to make quantitative surface topography measurements, and to construct composite images that maintain focus across large focal depths from several images spanning a range of focal positions. In a typical application, the sample is translated vertically in discrete increments relative to the objective lens. The appearance and subsequent disappearance of sample features provides an understanding of the three-dimensional nature of the sample.

Experimental Techniques

Microscope Configurations. The classification of confocal scanning optical microscopes is based on the light source and on the scanning technique used, as shown in Table 1. The two basic types use either laser or nonlaser illumination. The nonlaser configurations typically use high-intensity arc sources in conjunction with a Nipkow spinning disk for real-

time image scanning. Laser-based microscopes produce an image by scanning the beam across the sample in a TV-like raster pattern, and the data are stored, point by point, until a complete image area has been scanned.

Table 1 Hardware configurations for confocal microscopes

Microscope type	Acronym	Light source	Scanning technique
Tandem scanning reflected-light microscope	TSRLM, TSM	Nonlaser	Nipkow disk, separate illumination and image paths
Real-time scanning optical microscope	RSOM	Nonlaser	Nipkow disk, combined illumination and image paths
Laser confocal scanning microscope	LCSM	Laser	Dual galvanometer scanning of laser beam
Real-time laser confocal microscope	RLCM	Laser	Acousto-optic scanning of laser beam

In nonlaser confocal microscopes, emission from the broadband light source is collimated before illuminating a circular region of the Nipkow disk. Each aperture on the illuminated region of the disk forms one pinpoint of light (Fig. 4). Hundreds of points of light are produced simultaneously and focused on the sample. Reflected light is then refocused upon imaging apertures on the opposite side of the disk. Light that passes through the apertures forms a pattern of image points on the image plane. By spinning the disk at high speeds, the pinpoints of light scan across the sample to produce a full-color, live image. The image can be either visually observed or recorded with a video camera.

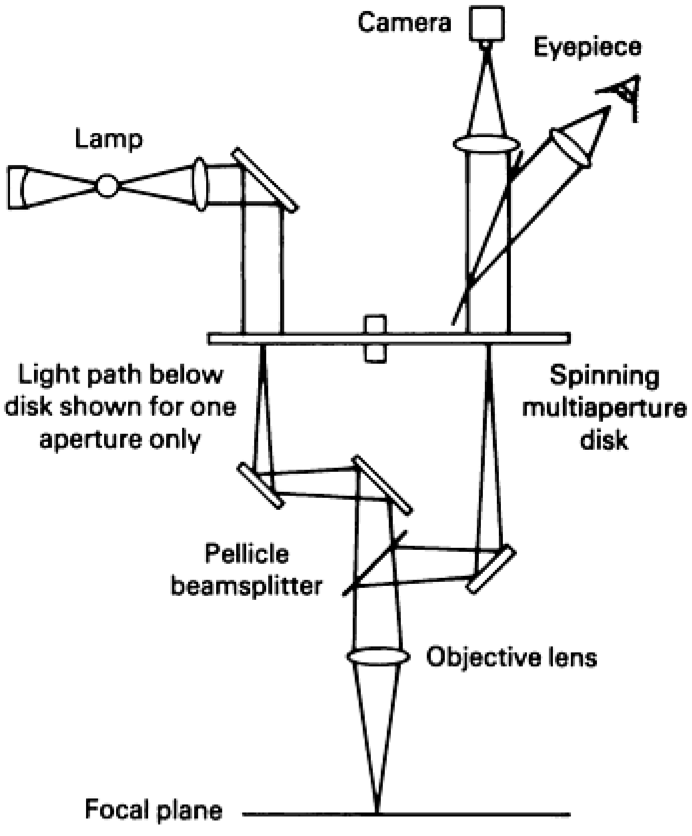


Fig. 4 Simplified optical diagram of confocal tandem-scanning reflected-light microscope

In a laser-based microscope, a single-wavelength laser beam is scanned to form a two-dimensional pattern of light pinpoints on the sample. Light reflected from the focal plane of the objective lens is focused upon a single aperture, and is then detected and stored to produce a two-dimensional image (Fig. 5). The most common scanning technique involves the use of dual galvanometers. However, a recently introduced variation involves the *x*-axis scanning of the laser beam at video rates, using acousto-optic deflection of the laser beam.

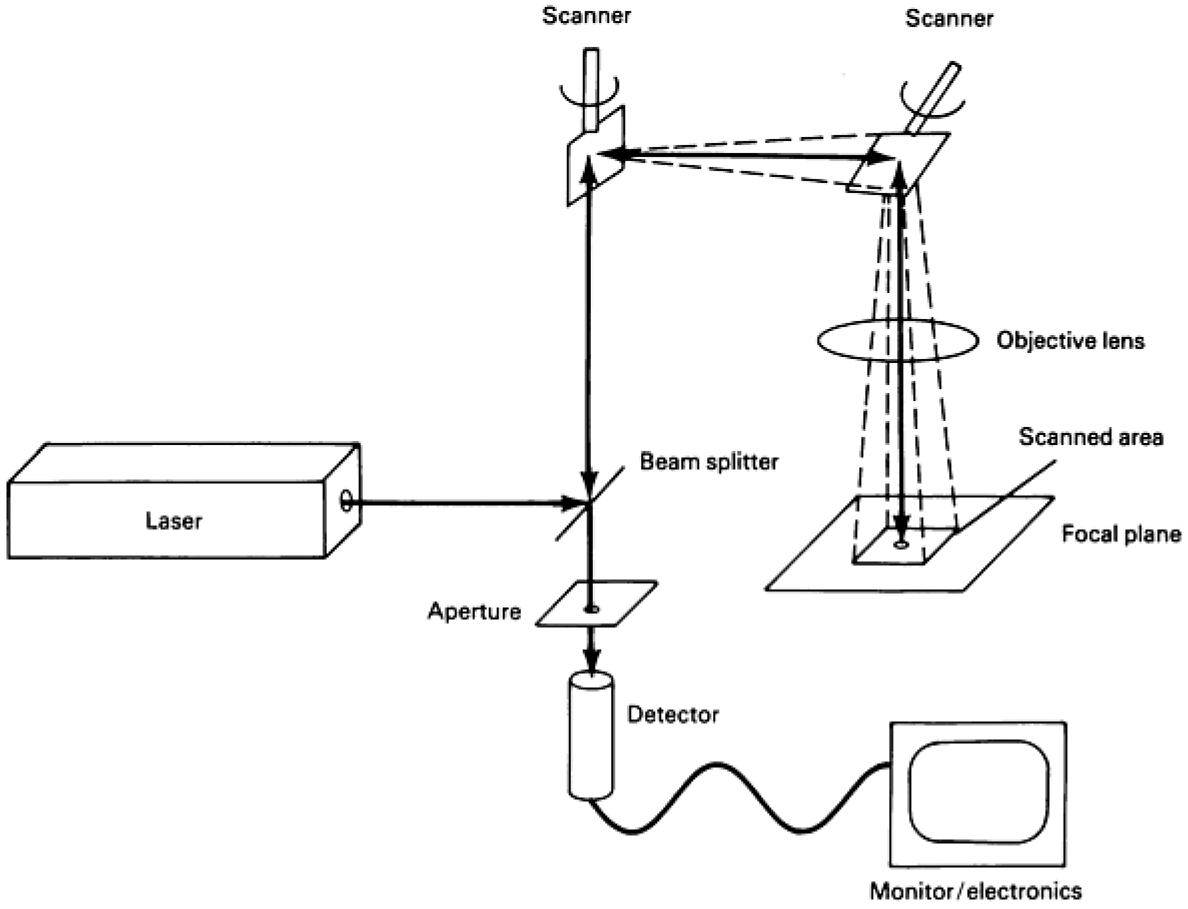


Fig. 5 Diagram of confocal laser scanning microscope

Other than the scanning technique, the primary differences between laser and nonlaser confocal microscopes involve the spectral content of the light, the brightness of the light beams, and the resolution. The nonlaser microscopes produce a full-spectrum image, rather than the monochromatic image inherent with laser illumination. For many samples, the full-color image provides many advantages that can be related to sample characteristics.

The intensity, inherent collimation, and monochromatic nature of laser illumination provide a brighter image, because the full power of the laser can be focused to a diffraction-limited spot. This spot brightness is particularly important in biological applications, where high intensities are needed to produce bright fluorescent images, but is not a significant advantage for reflective samples, such as those that are typical in metallography. The diffraction-limited spot of the laser beam also results in a 1.4× increase in the system lateral resolution.

Specimen Requirements and Limitations. Because confocal microscopy is an optical technique, specimen requirements are minimal. Samples do not need to be either electrically conductive or vacuum compatible. Opacity does determine which features of the sample can be studied, but both opaque and translucent samples can be analyzed. Although surface roughness can be measured for opaque samples, subsurface defects cannot be imaged.

Both surface and subsurface features of translucent samples can be observed, depending on the clarity of the material. In translucent specimens, the reflectivity and the tilt of the internal surface determine how well that surface will be visible in the optical section. Subsurface features that are overlaid vertically can create artifacts in computer-constructed images, and may increase the difficulty of quantitative measurements.

Two types of surface features are not suitable for analysis using confocal microscopy. Features in opaque samples that are obscured from line-of-sight view cannot be studied, because light cannot reach them. Features with nearly vertical sides, particularly cracks or pits where the depth is much larger than the width, are very difficult to measure. Because light from

these vertical features is often not reflected back to the objective lens for detection in the image, the features appear dark and out of focus.

Another limitation is the availability of objective lenses that have the optimum characteristics for viewing features of interest in the sample. Because the objective lens affects the three-dimensional resolution of the microscope, the availability of suitable lenses can limit the visibility of sample features. Critical objective lens characteristics include the working distance of the lens, and its field of view and numerical aperture. The former controls the magnitude of surface roughness that can be imaged clearly before the lens touches the sample, and the latter two control the maximum lateral size and vertical resolution limit of the features that can be analyzed.

Image Acquisition. The first step in the application of a confocal microscope is to survey the nature of a sample and determine which areas are most representative of the data being sought. Such surveys involve not only area analysis, but also depth analysis. Any existing pits, cracks, or fissures can be located and analyzed to determine whether a more detailed analysis of the area is appropriate.

The ease and simplicity of this process is one of the prime features of the tandem scanning reflected-light microscope (TSM) and the new real-time laser confocal microscope (RLCM) instruments. For some applications, such real-time analyses can be sufficient. However, for many applications, it is useful to further investigate the three-dimensionality of the sample and to extract such data in a usable form.

The next step is to acquire a series of images, known as a *z*-series, at specific increments across the total depth range of interest, and to store these images in either video or digital memories. Image capture--using a TV camera and image storage on videotape using VCR techniques--provides the fastest acquisition and lowest-cost storage method. However, the most powerful and common method is computer digitization of each acquired image as either a square or rectangular array of pixels. Each pixel typically has an intensity value that is based on 256 gray levels.

The *z*-series optical sections are typically acquired at precise vertical intervals. The proper selection of these intervals, which is based on the vertical resolution of the optical section, assures the operator that data are not missed (Fig. 6). The resulting images can be used for a broad range of sample analysis techniques that utilize computerized image processing.

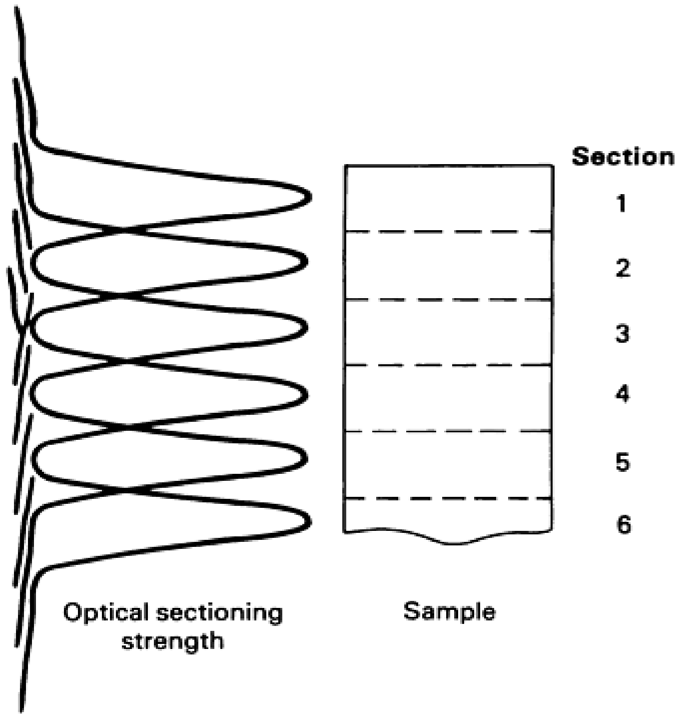


Fig. 6 Use of optical sections to completely cover entire sample depth with no missing information

Computerized Image Processing. Advanced image processing and analysis can be used with confocal microscopy to provide new techniques for studying surfaces. Both visual and quantitative data can be obtained by acquiring and processing a *z*-series of optical sections spaced in precise depth increments.

Table 2 lists image processing techniques that are used in confocal microscopy. For example, a composite through-focus image of a rough surface that maintains clear focus across the entire field can be generated. The final image comprises only the brightest (in-focus) pixels from all of the *z*-series images. Similarly, a topographic image can be produced in which the brightest pixel is gray-level coded so that it can be directly related to depth. The three-dimensional nature of these data not only allows the generation of three-dimensional images, stereo pairs, and isometric reconstructions, but makes it possible to perform profilometric and volumetric measurements. Such capabilities have proven particularly useful to researchers and technologists involved in surface morphology analyses.

Table 2 Image processing techniques for optical sections from confocal microscopy

Image type	Description	Pixel display mode
Through focus	Creates a clearly focused image with a large depth of field	Each pixel value in the final image equals the maximum intensity of any pixel at that location
Extended focus	Summation of multiple clearly focused images	Each pixel is a summation of all pixels at the same location
Topographic data	Display color or gray scale is proportional to the relative height of that feature on the surface of an opaque sample	Gray scale is proportional to the position of the optical section where maximum intensity value occurs
Topographic display	Projects the surface features in three dimensions using hidden line suppression techniques	
3-D projections	Contour lines of surface features are stacked in three dimensions for manipulations such as rotation, tilting, and cross-sectioning	
Stereo pairs	Two *z*-series are acquired by tilting the sample relative to the objective lens. Through-focus images are used to create an anaglyph image for viewing with stereo-optician-type optics	

Applications

Examples described in this section, which illustrate the application of confocal microscopy to characterize worn and damaged materials, include:

- The study of subsurface cracks that result from a scratch on the surface of a glass slide
- The analysis of arc-shaped cracks in the wear track of a SiC ceramic
- The evaluation of the topography of a wear track in a ceramic composite

Figure 7 shows a *z*-series of optical sections, spaced in increments of 5 μm (200 μin.). This series was produced by imaging below the surface of a scratch in a glass slide. The first optical section shows the scratch on the glass surface. The crack structure can be clearly observed in the succeeding optical sections as it initially descends vertically from the scratch, and then begins to branch out at 15 μm (600 μin.) below the surface. Notice that different features of the subsurface cracks come into focus at different elevations.

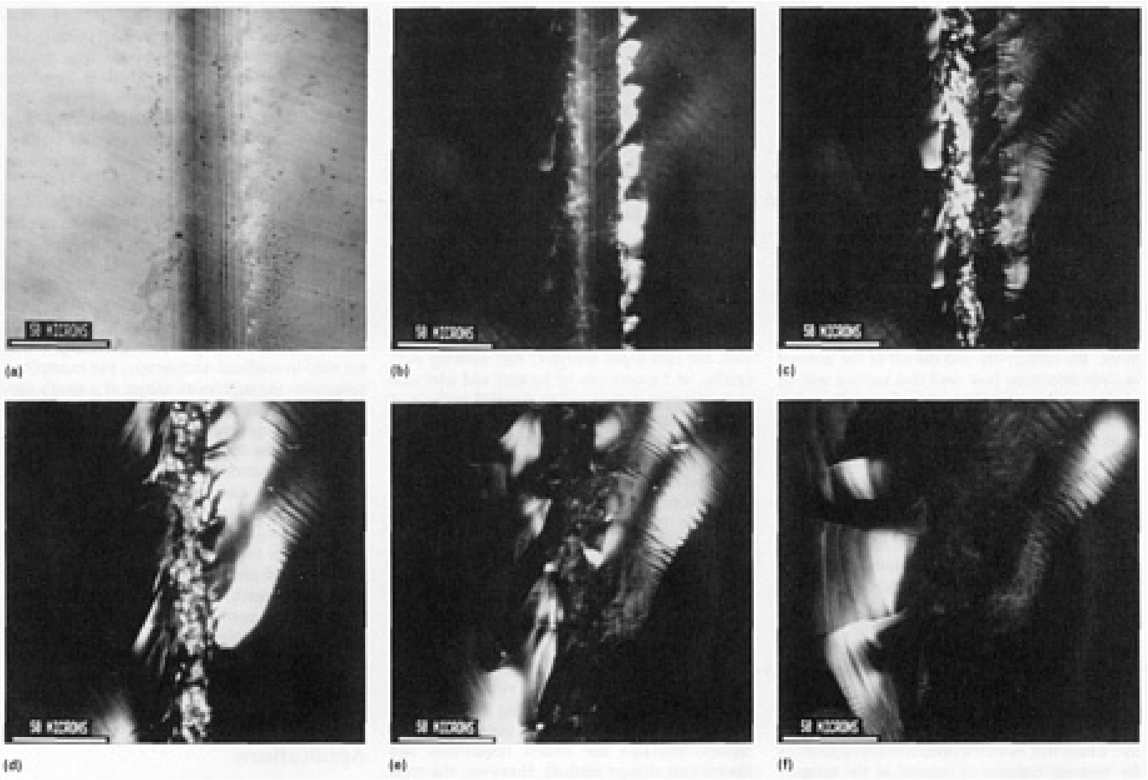


Fig. 7 Series of optical sections showing subsurface crack structure in glass slide, beginning at the surface in intervals of 5 μm (200 μin.). Scratch was made with a spherically tipped diamond stylus, 0.2 mm (8 mil) tip radius, using a 10 N (2.25 lbf) normal load and 13.0 mm/s (0.5 in./s) velocity.

Figure 8 shows the usefulness of through-focus and topographic images for analyzing the cracks produced in a wear experiment in which a SiC ceramic was tested. The through-focus image shows the entire field of view in the region of the cracks, maintaining clear focus. The corresponding topographic image displays the relative height of features using the intensity values in the image, so that the cracks show clearly in black. A linear analysis across the cracks can be used to analyze their depth, as well as width.

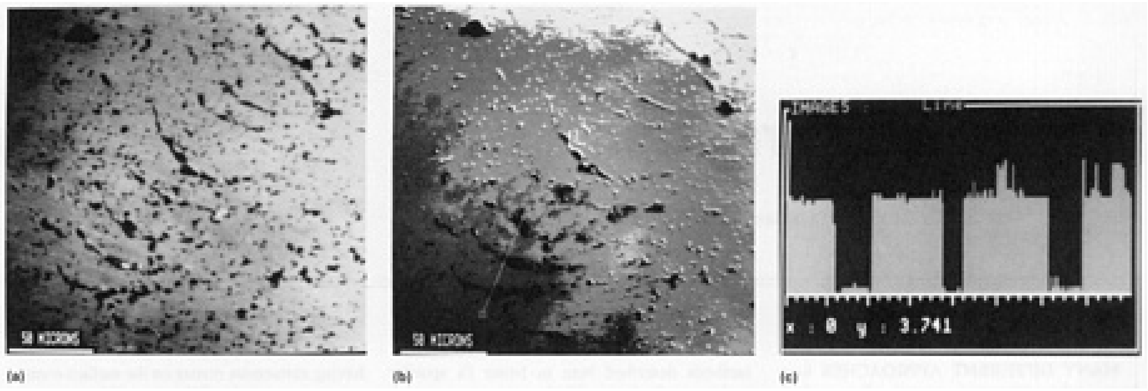


Fig. 8 Confocal microscopy analysis of a SiC ceramic wear specimen showing details of microcracks. Wear damage was produced by a silicon nitride ball sliding unlubricated over the surface with a 10 N (2.25 lbf) load and 0.1 m/s (4 in./s) sliding velocity in air at room temperature. (a) Through-focus image. (b) Topographic image. (c) Linear analysis along line displayed in (b), with a vertical range of 3.1 μm (125 μin.)

Figure 9 shows how the larger topography of a wear track can be evaluated and then displayed as a three-dimensional projection. Again, the topographic image was used to illustrate surface roughness, and linear analysis was used to quantitatively analyze the wear track. The image was then projected into three dimensions and tilted to reconstruct the surface, providing more visual details. This image is more intuitive to interpret than its source topographic image.

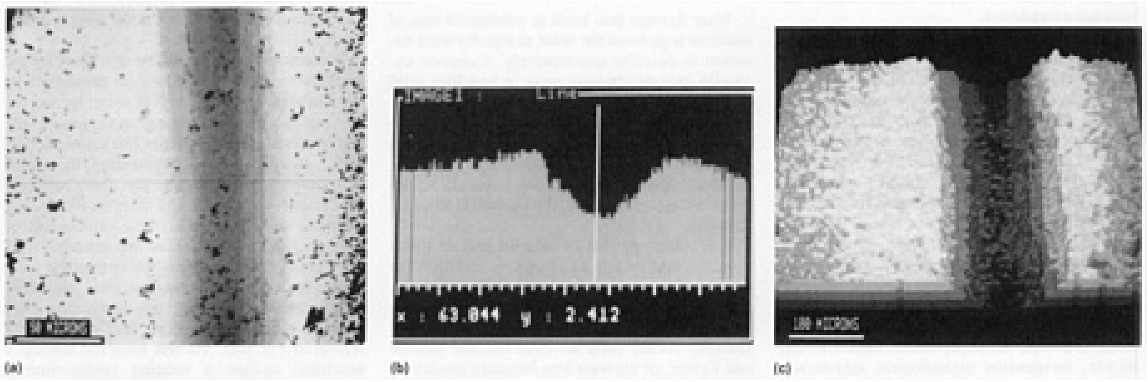


Fig. 9 Confocal microscopy provides topographic analysis of wear track. (a) Topographic image. (b) Linear analysis along line displayed in (a), with vertical range of 4.9 μm (196 μin.). (c) Topographic display using hidden line suppression techniques

References

1. M. Richardson, Confocal Microscopy and 3-D Visualization, *Am. Lab.,* Nov 1990, p 19-24
2. G. Kino and T. Corle, Confocal Scanning Optical Microscopy, *Phys. Today,* Sept 1989, p 55-62
3. B. Yatchmenoff and R. Compton, Ceramic Surface Analysis Using Optical Sections, *Ceram. Bull.,* Vol 69 (No. 8), 1990, p 1307-1310
4. M. Minsky, U.S. patent 3013467, 1957
5. M. Petran, M. Hadravsky, D. Egger, and R. Galambos, Tandem-Scanning Reflected-Light Microscope, *J. Opt. Soc. Am.,* Vol 58, 1968, p 661-664
6. P. Nipkow, German patent 30,105, 1884
7. N. Aslund, A. Liljeborg, P.-O. Forsgren, and S. Wahlsten, Three-Dimensional Digital Microscopy Using the PHIOBOS Scanner, *Scanning,* Vol 9, 1987, p 227-235
8. S. Inoue, Foundations of Confocal Scanned Imaging in Light Microscopy, *Handbook of Biological Confocal Microscopy,* J.B. Pawley, Ed., Plenum Press, 1990, p 1-14

Wear Measurement

A.W. Ruff, National Institute of Standards and Technology

Introduction

MANY DIFFERENT APPROACHES have been used to determine amounts of wear, both qualitatively and quantitatively. An examination of the technical literature (Ref 1 and 2, for example) and published standards (Ref 3) that relate to wear reveals the breadth of techniques that have been employed.

There are a number of reasons as to why this situation exists. One is that the type of wear damage can take several forms. If material is actually lost during wear, then a commonly used method of measurement is to determine the amount of removed material, perhaps by weight loss, as is also done in the field of corrosion. Alternatively, if the wear process leads to surface distress on some component, then surface roughening or cracking may be measured. Other forms of surface and subsurface wear damage can be encountered, as well, and can be measured by other direct and indirect methods.

The amount of wear will also influence the selection of measurement method. If large amounts of wear are experienced, then relatively simple, inexpensive measurement approaches, such as volume change or mass change determination, are usually conducted successfully. Alternatively, if very small wear amounts are experienced, then more sensitive and costly techniques are necessary to detect minute changes of mass or volume.

The type of investigation being conducted is yet another factor that influences the choice of wear measurement method. In a research laboratory, it may be essential to carry out a highly precise measurement in order to correctly identify differences among either the effects of research parameters or the materials involved. In contrast, if a field study is being conducted to determine the basis for proposed changes in usage conditions for a tribosystem, then perhaps less exacting measurement options would suffice as the basis for a valid decision.

An objective of this Volume is to provide examples of problems and solutions that are encountered in the wear field. Therefore, this article describes many of the most common methods used to measure wear, based on the literature, and presents the information in sufficient detail for the reader to make a selection for a problem at hand. It is certainly appropriate, and in some instances necessary, to modify the methods described here to better fit specific problems. Such modified wear measurement methods add to the wealth of approaches available in this field.

Mass Loss Measures of Wear

Wear damage that leads to substantial loss of material is perhaps the most straightforward situation to describe quantitatively. Common examples include abrasive wear in handling solid materials, as occurs in the mining industry. Wear loss can be determined by measuring either mass change or dimensional change, which is discussed in the next section of this article. Because most laboratories have access to equipment that weighs objects, the method is straightforward.

It is necessary that an original part or specimen (or equivalent) be weighed, and that the weight of the object after wear exposure be determined and subtracted from the original to determine the difference in weight (that is, mass change). As the parts involved become smaller and lighter, or the wear loss becomes smaller, it will be necessary to use increasingly sensitive weighing equipment. At some point, the mass change will be too small for the method to be feasible.

Other problems with this approach include the need to clean the specimen carefully to avoid having extraneous matter on the surface contribute to any weight difference. Of course, any fluids or solids used in cleaning must be thoroughly removed or dried. Another consideration is that material that was plastically displaced by the wear process but not actually removed from the part will not be included in the weight difference.

The amount of wear can be described by the absolute amount of mass loss (in grams), or by the rate of mass loss per unit of usage (grams per day), or by a fractional change in the mass of the part involved (1% change per 100 hours of operation). In many areas of engineering, the choice of reporting unit is frequently a conventional one. In most of the ASTM wear standards, the reporting unit of wear is cubic millimeters of volume, rather than mass, so that materials with different densities can be better compared.

One example in this category, from laboratory measurements, could be the application of a standard abrasive wear test, ASTM G 65. As shown in Fig. 1(a), the test involves loading a specimen against a rotating rubber-rimmed wheel while a flow of abrasive sand is directed at the contact zone. This test is widely used by industry to assist in selecting materials for abrasive wear service. Choices of loads and sliding distances are detailed in the test method. A photograph of a wear scar on a typical specimen is shown in Fig. 1(b). Because the test is a standard, it has been used in many interlaboratory studies of different materials.

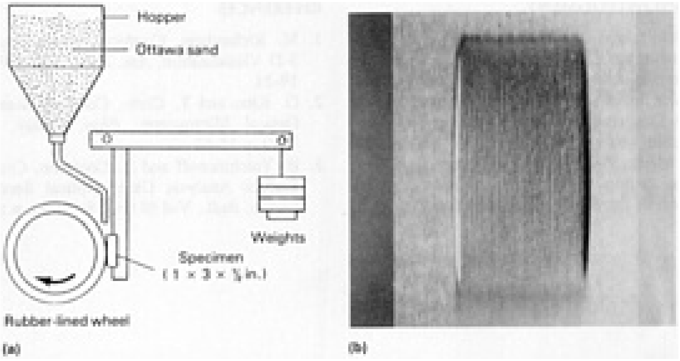


Fig. 1 (a) Schematic of standard abrasive wear test, ASTM G 65. (b) Wear scar on typical specimen

One set of results from a National Institute of Standards and Technology (NIST) study of a tool steel, shown in Fig. 2, is expressed in terms of mass loss (in grams) and wear scar maximum depth (in inches). Some idea of the observed variation from test to test can be seen. Intralab precision for this test can be as low as 3% (relative), and interlab precision, as low as 5%. Additional data on other materials can be found in the standard itself. The approach of measuring mass change in this test method is usually quick and inexpensive, and specimen costs can be low. Weld-overlay materials, coatings, ceramics, composites, and many other types of materials can be studied using this method.

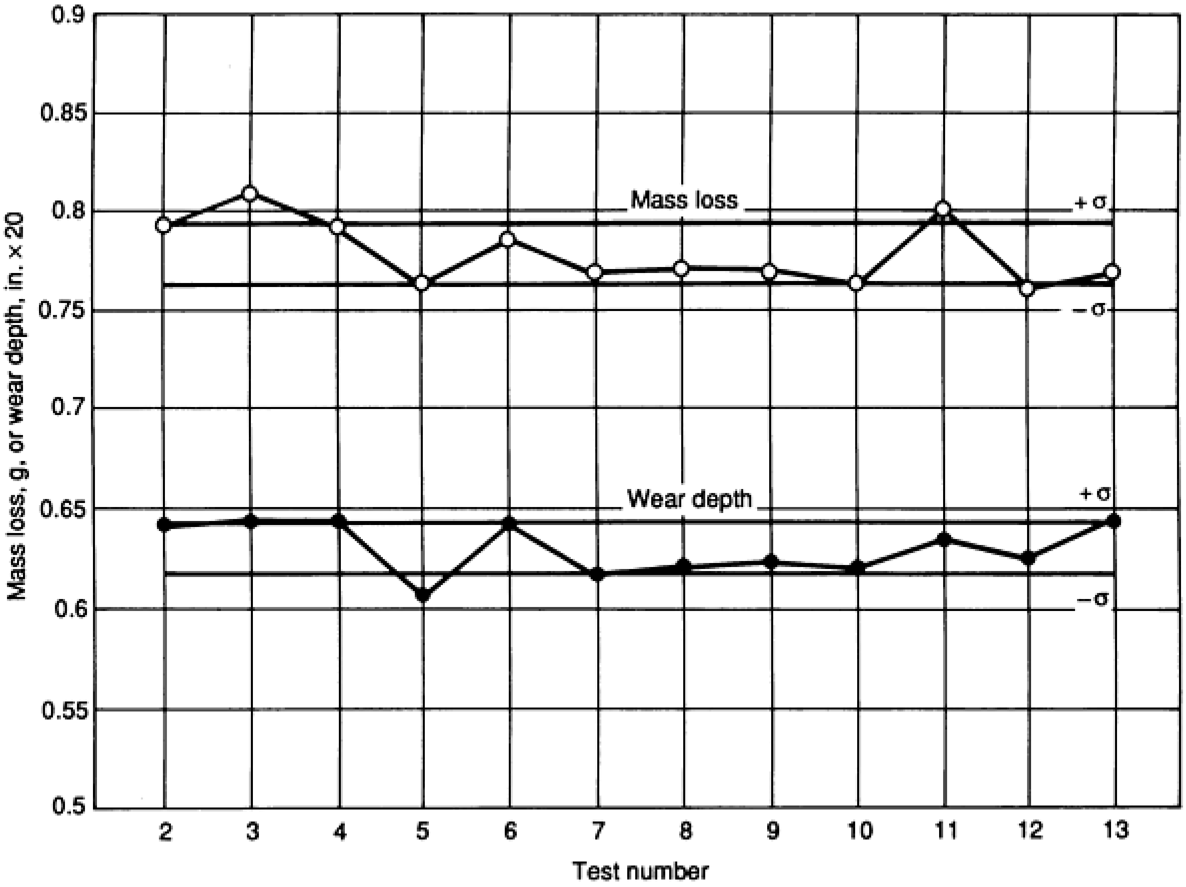


Fig. 2 Wear loss results (from a NIST study of a tool steel) expressed as mass loss (g) and wear scar maximum depth (in.). σ**, standard deviation**

Another example in this category is drawn from a report of a study of coins in circulation (Ref 4). The purpose of the study was to examine the serviceability of different coin materials, as determined by loss of mass during circulation. Figure 3 shows the results of sampling a particular denomination of coin removed from circulation. Although the variation in mass for any particular year is large, the trend of increasing wear with circulation time is clear. Because the service conditions were not controlled, the scatter in the wear data is not surprising.

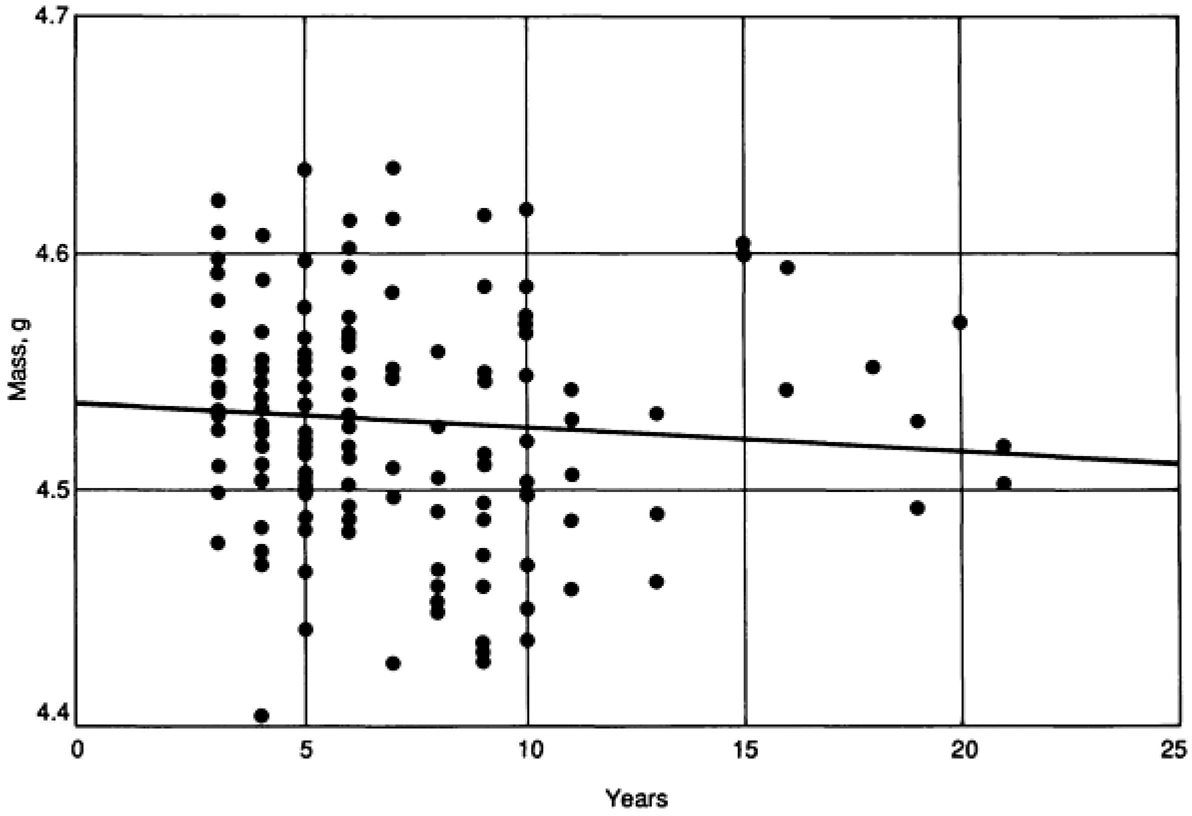


Fig. 3 Wear loss results for particular denomination of coin removed from circulation, showing a trend of increasing wear with circulation time, with considerable scatter of the data. Source: Ref 4

To simulate use conditions, a laboratory test involving tumbling of coins inside of slowly rotating cylinders under controlled conditions was developed. The tests showed the importance of surface hardness in reducing wear rate, and the influence of test parameters (for example, cylinder diameter and number of coins tested together) on the results (Fig. 4). In the laboratory study, weight loss was used as the measure of wear, and a sensitivity of 0.01% was achieved.

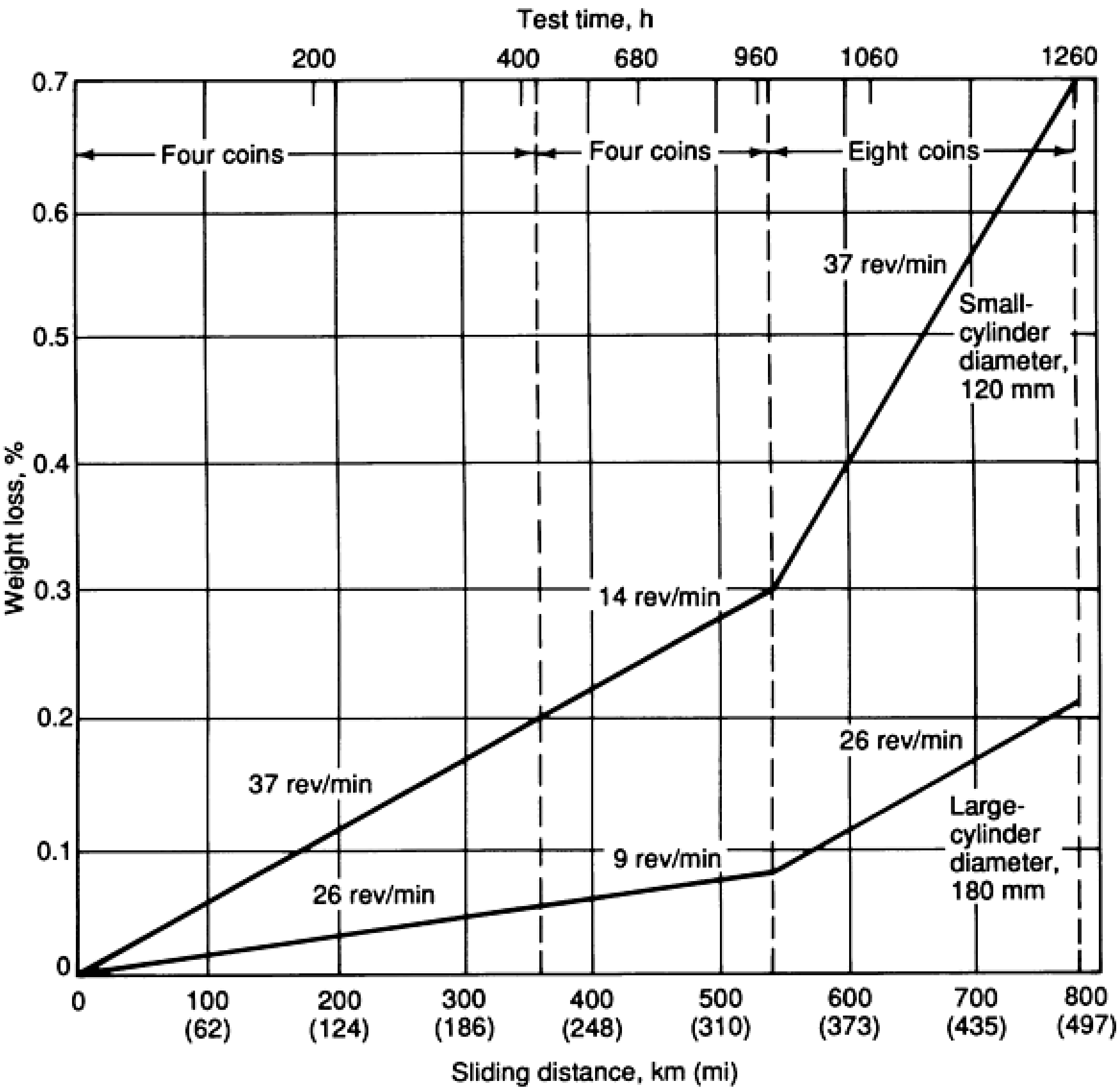


Fig. 4 Wear results from laboratory test involving tumbling of coins under controlled conditions to simulate use conditions. Source: Ref 4

Linear Measures of Wear

A common alternative to the weight loss measure of wear is to measure dimensional change. In many situations, the design of a component that is subject to wear will only allow up to a certain loss of dimension before either the integrity or function of the system is lost. In such cases, monitoring the dimensions of a part is a natural approach to assessing the amount of wear encountered. Frequently, such studies lead to the establishment of criteria for servicing or for part replacement. Examples include the wear of bushings or shafts, ball-bearing retainers, sliding actuator parts, and piston-cylinder wall contacts. Although this approach to wear measurement is frequently used, very small amounts of wear are difficult to measure, much as in the case of wear measurement by mass loss.

As one example of a linear measure of wear, consider a reported study of the abrasive wear of various types of metals used in tractor shovels (Ref 5). In the design of interest, both bucket teeth and cutting edges were involved. Therefore, a natural approach was to measure the shape changes of those components that were due to wear. Figure 5 shows a worn bucket tooth having a linear wear loss of amount *L*. Because that measure only applies to one location on the worn part, the user must be certain that the location picked is either typical or average, or perhaps shows a maximum amount of wear.

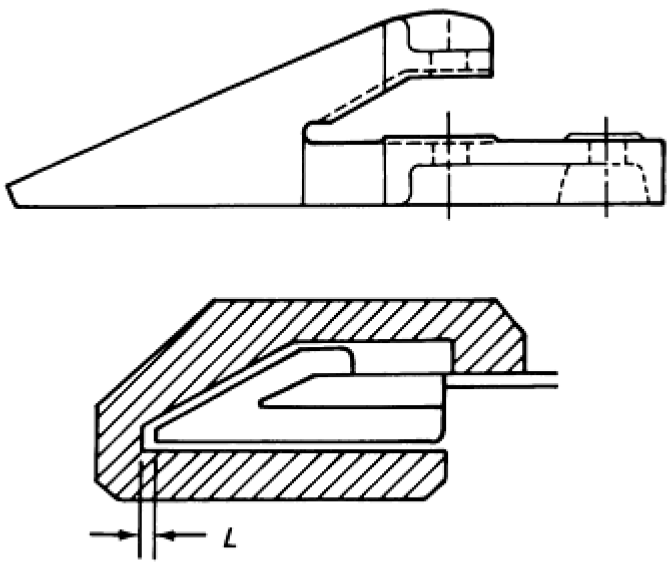


Fig. 5 Worn bucket tooth showing linear wear loss of amount *L*. Source: Ref 5

Several field test results on different bucket teeth are shown in Fig. 6. Because the wear lengths involved in this example are substantial, the precision of this measurement is not of concern. Results of this type can be directly converted into life limits (of wear) for the materials and designs of interest.

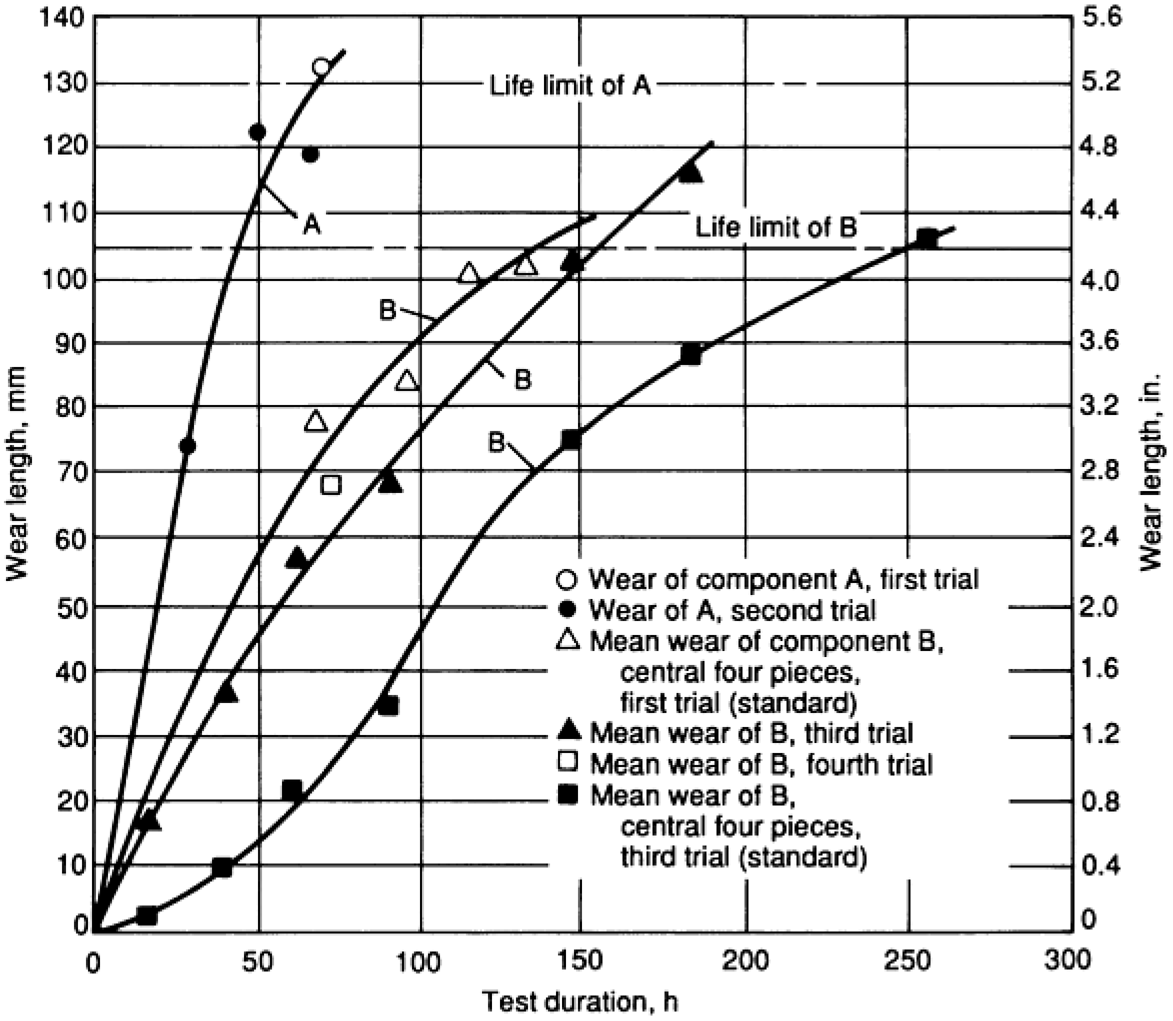


Fig. 6 Wear length results from different components and tests (Ref 5)

A second example of linear wear measures is drawn from laboratory testing results obtained using a standard test methodology. The standard, ASTM G 99, describes a method for pin-on-disk sliding wear testing (Ref 6). The pin, which in many cases shows the majority of wear for the two specimens in contact, is typically one of two shapes: a small solid cylinder with a spherical end or a sphere that is constrained not to rotate while in sliding contact. In either case, a wear scar develops during the test. The scar is relatively flat and circular in cases where the disk wear is relatively small. Figure 7 shows a steel pin wear scar taken from an interlaboratory test series (Ref 7).

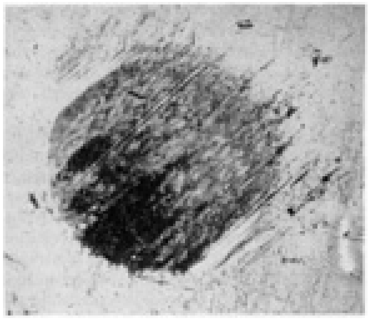


Fig. 7 Wear scar on spherically surfaced steel pin taken from test that was part of interlaboratory comparison. Magnification: 93×

Figure 8 presents a set of scar measurements from five different laboratories in the test series mentioned above. The graph indicates the variability found within and between the different laboratories. The average scar diameter was 2.13 mm (0.09 in.) with a standard deviation of 0.35 mm (0.014 in.). Those measurements involve optical microscope determinations of the average wear scar diameter, an approach that can usually be carried out to the nearest 0.01 mm (0.4 mil). One advantage of this particular wear test specimen configuration is the ease of measuring small amounts of wear.

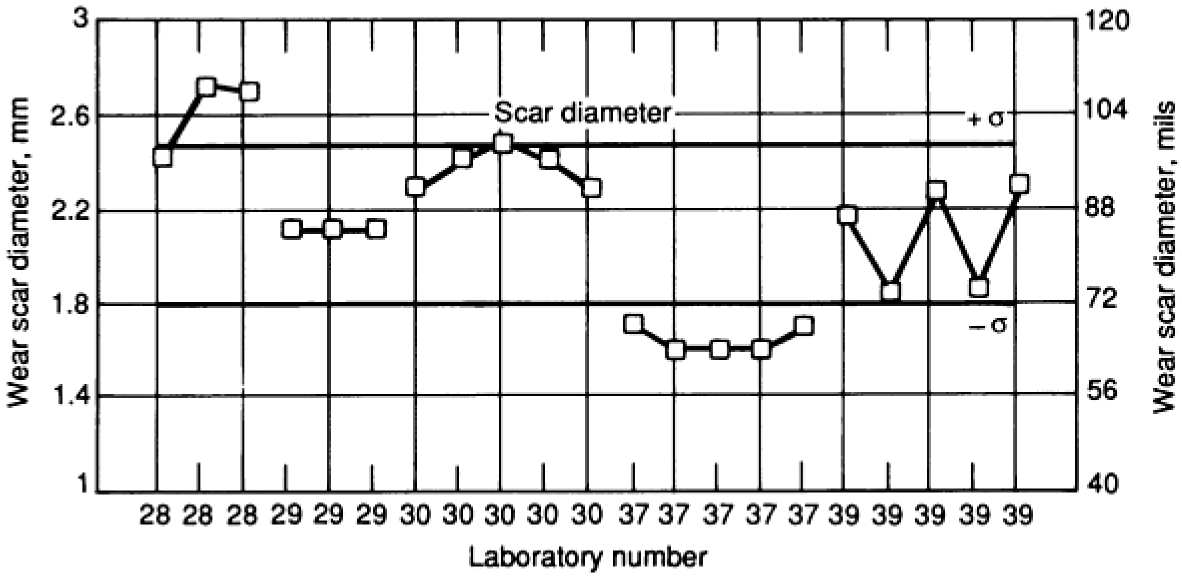
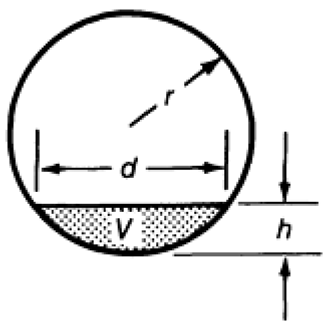


Fig. 8 Set of wear scar measurements from five different laboratories; graph indicates the variability found within and between the different laboratories. σ, standard deviation

Although the basic method involves a linear measurement of wear amount, in many cases the result is converted to a calculated wear volume, usually for the purpose of comparing materials with different densities. That conversion uses geometric relations and assumes a planar wear scar; the equations are shown in Fig. 9 (Ref 8) and discussed in ASTM standard G 99. In some cases, a flat-ended pin is used in this type of test, and then the change in pin length can be determined and used as the wear measure, although with considerably reduced sensitivity.



d = wear scar diameter
V = worn volume

$$V = \frac{\pi}{6}h\left(3\frac{d^2}{4} + h^2\right) = \frac{1}{3}\pi h^2(3r - h)$$

$$h = r - \sqrt{r^2 - d^2/4}$$

For $h/r \ll 1$

$$h \approx \frac{d^2}{8r}$$

$$V = \frac{\pi d^4}{64r}$$

Fig. 9 Geometric relations used to convert wear scar size to wear volume, assuming a planar wear scar on a spherical specimen; see ASTM Standard G 99 for discussion. Source: Ref 8

A novel use of a linear measure of wear was developed by Glaeser (Ref 9), who placed microhardness indentations on the surface of a bronze bushing and measured their size changes during the running time of the bushing as an indicator of wear. A variant of that method (Ref 10) involved placing a number of indents of different depths on a wearing surface, and periodically determining which indents remained after various amounts of wear occurred on the surface. Scratches of different controlled depths can also be used in this way.

Area Measures of Wear

Certain wear contact geometries produce material loss over a localized area on the two surfaces. In many cases, those areas of wear loss can be measured and are proportional to the amount of wear. Examples would include worn areas on gear teeth, on bearing retainers, and on sliding pads with contoured surfaces. If the curvature of the surface is known, then the amount of wear can be quantified on the basis of the area worn. Because many tribological components involve area contacts, as contrasted to point or line contacts, area measures of wear are important. Two examples are described below.

One frequently used laboratory test system comprises a stationary block and a rotating ring. Several ASTM standards, such as Ref 11, that are concerned with lubricants and material wear, utilize this type of system. Although the initial contact between the two specimens is nominally a line (there is actually a small lateral width associated with elastic deformation along the contact line), the resulting scar on the block becomes a curved rectangular surface as the two components wear. The volume worn from the block can be calculated from the two scar dimensions and the ring (or scar) curvature, but it is also common to find the projected scar area reported. The ASTM standard involves scar width measurement only because it specifies the specific block size, and, hence, scar length and ring diameter.

An optical micrograph of a portion of a wear scar on a Ag-Cu alloy block is shown in Fig. 10, along with a profile measurement taken perpendicular to the scar length (that is, the sliding direction). Figure 11 shows a schematic of the

worn volume in this type of geometry and the geometric relations that are used to calculate wear volume under the assumption of an ideal scar shape.

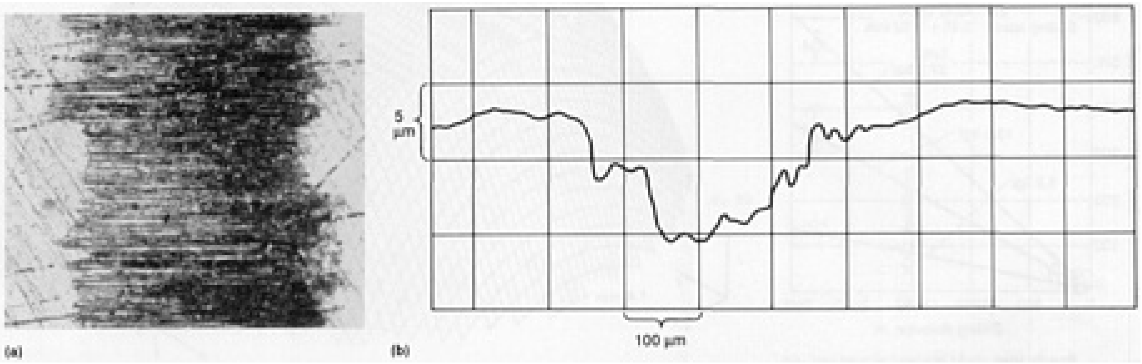
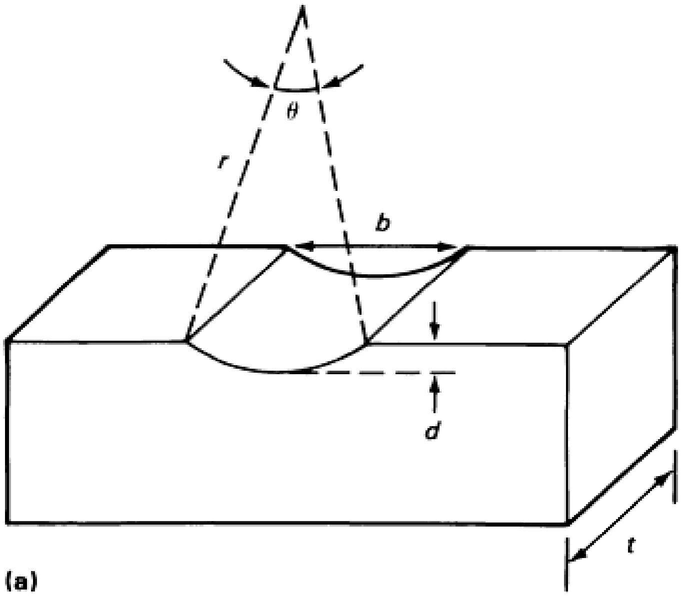


Fig. 10 (a) Optical photograph of portion of wear scar on a Ag-Cu alloy block. Magnification: 32×. (b) Profile measurement taken perpendicular to scar length (in sliding direction)



(a)

Scar width $= b = D \sin \frac{\theta}{2}$

Scar volume $= \frac{D^2 t}{8} (\theta - \sin \theta)$

where $\theta = 2 \sin^{-1} \frac{b}{D}$

Scar volume $= \frac{D^2 t}{8} \left[2 \sin^{-1} \frac{b}{D} - \sin\left(2 \sin^{-1} \frac{b}{D}\right) \right]$

(b)

Fig. 11 (a) Worn volume of block specimen (top), where *t* is block width in mm, *r* is radius of ring in mm, *D* is 2*r* (diameter of ring) in mm, *b* is average scar width in mm, θ is sector angle in radians, and *d* is scar depth in mm. (b) Geometric relations used to calculate wear volume, assuming ideal scar shape. Source: Ref 11

An example of a laboratory study of wear that utilized a block and ring test method involved ceramic materials (Ref 12). The role of iron oxides, which were present on the surface of a rotating steel ring, in causing sliding wear of blocks of aluminum oxide was examined. Wear volume was calculated based on actual measurements of the scar width and length, and on the assumption that the scar surface curvature matched that of an unworn ring. Examples of the results, which indicate the effect of load and sliding distance on wear of the ceramic block, are shown in Fig. 12.

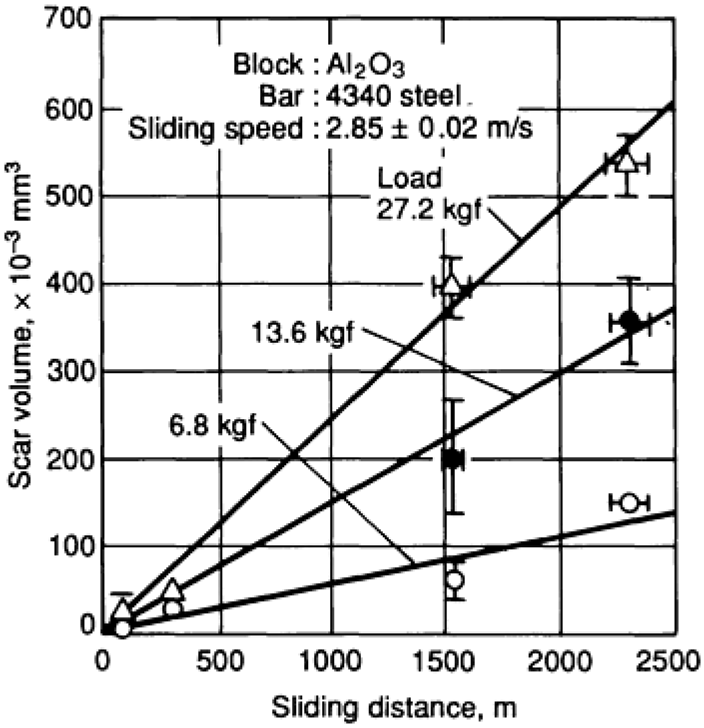


Fig. 12 Results from study of effect of load and sliding distance on wear of aluminum oxide test blocks versus steel. Source: Ref 12

The actual surface curvature within a worn area is often a significant point that is not carefully addressed. Earlier, it was stated that wear on a spherically curved surface can be expressed by a single measurement of circular scar size. However, that too involves an assumption about the flatness of the scar region. In general, it is necessary to determine the surface profile associated with the worn region. Based on those measurements, approximate methods for calculating wear may or may not be justified.

Another common laboratory wear test geometry requires the use of surface profiles to quantify the wear amount. This geometry involves reciprocal sliding of one body against a flat surface, usually referred to as a pin-on-flat test system. Wear on the flat specimen can be measured by determining the surface profile in the wear scar, perpendicular to the sliding direction. Numerical integration of a representative set of the profile curves gives the cross-sectional area of the scar and leads to an amount of wear per unit length of the sliding contact. One example of a wear scar profile from a wear test block was previously given in Fig. 10.

A portion of a linear wear scar on a flat (high-purity iron) specimen is shown in Fig. 13, along with two examples of the cross-sectional surface profiles at different locations. The wear cross-sectional area can be determined from such profiles. In the example shown, plastic deformation in the flat specimen has led to ridges of material on either side of the wear scar; such displaced material is usually ignored in calculating the wear cross-sectional area.

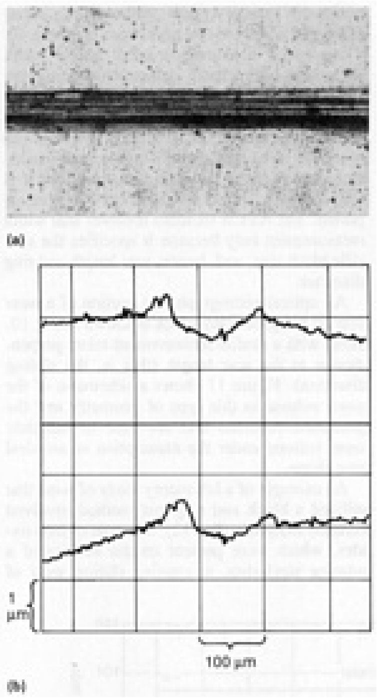


Fig. 13 (a) Portion of wear scar on flat, high-purity iron specimen tested in reciprocal sliding; magnification: 45×. (b) Two examples of cross-sectional surface profiles of scar at different locations, which can be used to measure wear area

Volume Measures of Wear

In a large proportion of reports that contain wear measurements, one finds the wear amount reported in volume units, for example, mm^3. This better enables a comparison of wear among materials having different densities, and also permits easy calculation of linear wear amounts or wear allowances. Many of the standards for wear testing require the reporting of wear in volume units (Ref 3).

In some cases, it is actually necessary to directly measure wear volume. This generally occurs when the worn region is very irregular or unsymmetric in shape, or when high accuracy in the result is needed. Unfortunately, such measurements are quite time-consuming. Two examples of direct volume determinations associated with laboratory wear testing are given next to illustrate the methods involved.

In one study (Ref 13), the worn surface of interest was neither flat nor smooth enough to permit accurate use of the usual geometric formulae. Therefore, the surface was traced on an *X*-*Y* stylus profiling system, and the resulting data were digitized to facilitate further calculations. Figure 14 shows a view of the worn surface reconstructed from the surface profile data. With such data, it is straightforward to calculate the volume difference between the worn specimen and the unworn original, which in this case was a sphere. This approach should be possible for most worn contacts, as long as they can be cleaned of extraneous matter, such as wear debris particles, and as long as sufficient lateral and vertical resolution are offered by the stylus tip and system.

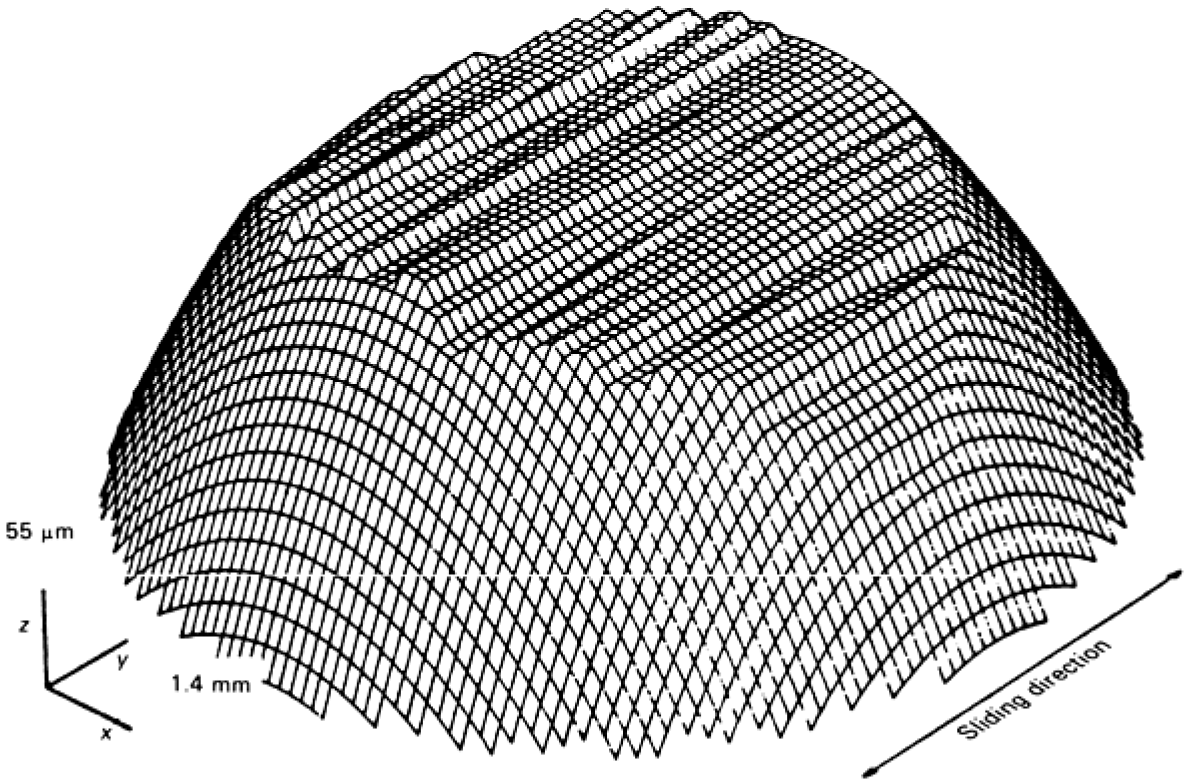


Fig. 14 Worn spherical surface reconstructed from three-dimensional surface profile data. Note that magnifications in *X* and *Y* directions are both the same, but are much less than in *Z* direction.

Another example focuses on the phenomenon of galling wear. When certain material combinations are moved relative to each other in contact under high loads, it is possible to cause localized material adhesion. This process of local seizure leads to severe damage on the moving surfaces (Ref 14). The damage consists of depressions in either surface where material has been torn out, and protrusions on either surface where material has been added, by a process sometimes described as cold-welding. It is possible to describe such damage in quantitative terms in several ways (Ref 15).

One approach is to measure the volume of material either removed from or added to the original surfaces, based on three-dimensional profiling of the contact regions on the surfaces. A reconstruction of one of the surfaces after a single sliding pass is shown in Fig. 15 in oblique projection using the profile data.

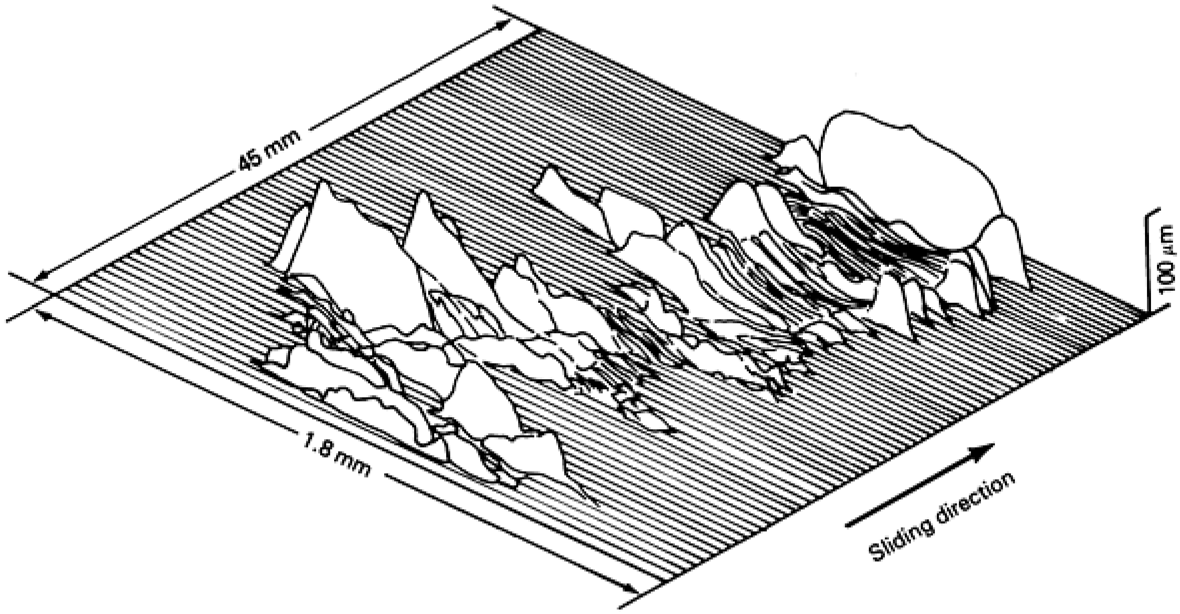


Fig. 15 Reconstruction of galled surface after a single sliding pass shown in oblique projection using three-dimensional profile data. Source: Ref 15

The worn or displaced volume can be calculated by numerical methods from the profile data; those results are shown in Fig. 16 for a nickel flat specimen (the opposing surface was copper). The volume shown represents incremental slices perpendicular to the sliding direction at different distances of sliding. At the initial part of the wear track, considerable copper material was deposited, and the amount decreased as sliding continued. Further details on the method can be found in Ref 15.

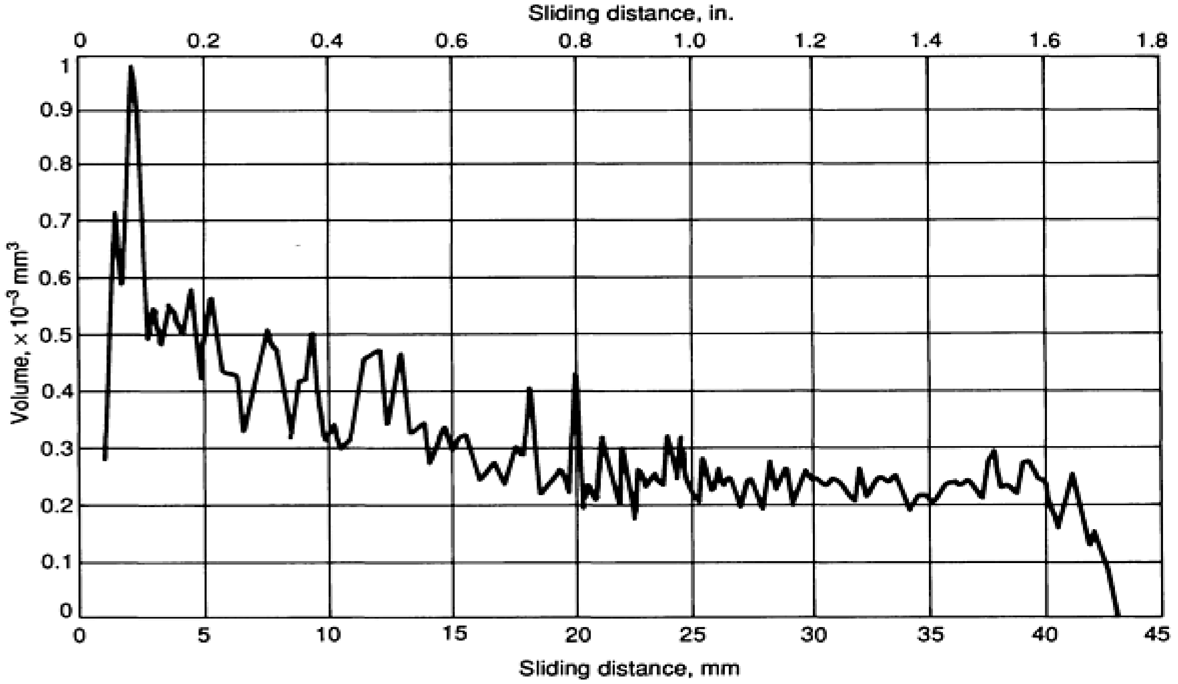


Fig. 16 Results of calculation of worn (displaced) volume for a nickel flat specimen. Volume shown represents incremental slices perpendicular to sliding direction at different distances of sliding.

Other Measures of Wear Damage

Wear is defined in ASTM G 40 as: "Damage to a solid surface, generally involving progressive loss of material.. . ." This statement recognizes the practical importance of surface damage and surface alteration in affecting, possibly, the function of a tribological component, particularly for systems designed with close tolerances in critical areas. The discussion that follows is intended to alert the reader to several other important issues often associated with the wearing process.

In many mechanical systems, certain contact surfaces are manufactured with close control over surface roughness and finish. They can be surfaces that are lubricated by oil during operation, requiring a surface finish of suitable smoothness to maintain an adequate but thin lubricant film in the contact zones. During service, those surfaces may become roughened and scratched, leading to regions where the lubricant film may not be able to cover the peaks in the topography. Such regions are then prone to local adhesion and seizure. In such a case, measurement of the state of wear would involve visual inspection of the surface and measurement of the current value of surface roughness. Published failure analyses in the technical literature give evidence that changes in surface finish are a frequent factor in tribological component failure. Often, scratches are found to have occurred during either installation of a part or servicing of a system. Abrasive particle contamination, sometimes involving lubricant contamination, is also a frequent source of surface scratching. Microscopic analysis and profile measurement of the contact surface are valuable techniques that reveal damage and change in surface finish.

A related issue is surface deformation during service. In component designs that involve high contact pressures, plastic flow of material may occur in the contact region over a period of time, altering surface topography. Principal causes of surface deformation are higher loads than the design allowed and lower resistance to plastic flow in certain regions of the material, possibly because of insufficient processing control on composition, microstructure, and hardening treatments. Microscopic examination of the contact surface may reveal evidence of surface deformation.

In certain types of materials, resistance to localized cracking (toughness) can be relatively low, and the designer involved must carefully consider the possibility of local fracture processes that are associated with the contact region and lead to component failure. Examples of such materials are ceramics, as well as metals that are either treated to high levels of hardness or possibly used at very low service temperatures. Local surface cracking may develop slowly, with time, as a fatigue process, or even suddenly, as a result of a transient overload on the contact. In any event, the cracks that develop may continue to grow with time and eventually lead to loss of material, either in the form of surface spalls or by crater formation. Surface examination can reveal the cracking, although in many cases the cracks may be obscured by other surface topography. Dye penetrant treatment and examination may be a useful approach for detection of small cracks.

An important but poorly appreciated feature of tribological contacts is the pattern of stresses in and beneath the contact surfaces. Frequently, the maximum shear and tensile stresses are found somewhat beneath the contact surface. The depth depends on material properties, load, and friction coefficient. This situation can lead to subsurface damage of one form or another. Thus, an observed lack of any component surface damage during service cannot be taken as an indication of the true subsurface condition. Cracking can be an active subsurface process, particularly near particulate inclusions, with little surface evidence of the problem. The detection and analysis of such cracks may be difficult; possible techniques to apply include x-radiography and ultrasonic inspection or destructive cross-section metallography.

Another common feature in wearing systems is material transfer. The earlier discussion of galling wear provided an example. Much less severe transfer can also lead to failure problems. In unlubricated sliding systems, it is very common to find transfer films that are formed on the contact surfaces. The films represent an intermediate stage in the wear process and are a source of wear debris as they change dynamically during service. Both the friction and wear characteristics of the system can be largely determined by the characteristics of such transfer films.

In lubricated systems, other types of films, both desirable and undesirable, frequently form. In the former case, most lubricants contain chemical additives that are designed to deposit very thin films of reaction products in order to aid in surface protection and to lower friction. In the latter case, thicker films can form from the service (process) environment and interfere with the tribological function. Surface inspection and, possibly, chemical microanalysis may be needed to determine the nature of surface films that are present.

This brief discussion of other attributes of tribological contact surfaces is not intended to completely describe all possibilities. Each case of interest must be evaluated in the context of the design requirements and the tolerance for surface change and damage.

Calculation of Geometric Wear Volume

The equations shown in Fig. 9 and 17 can be used to calculate the volume of material removed from a workpiece of specific geometry due to wear. These particular shapes are frequently used in laboratory wear testing.

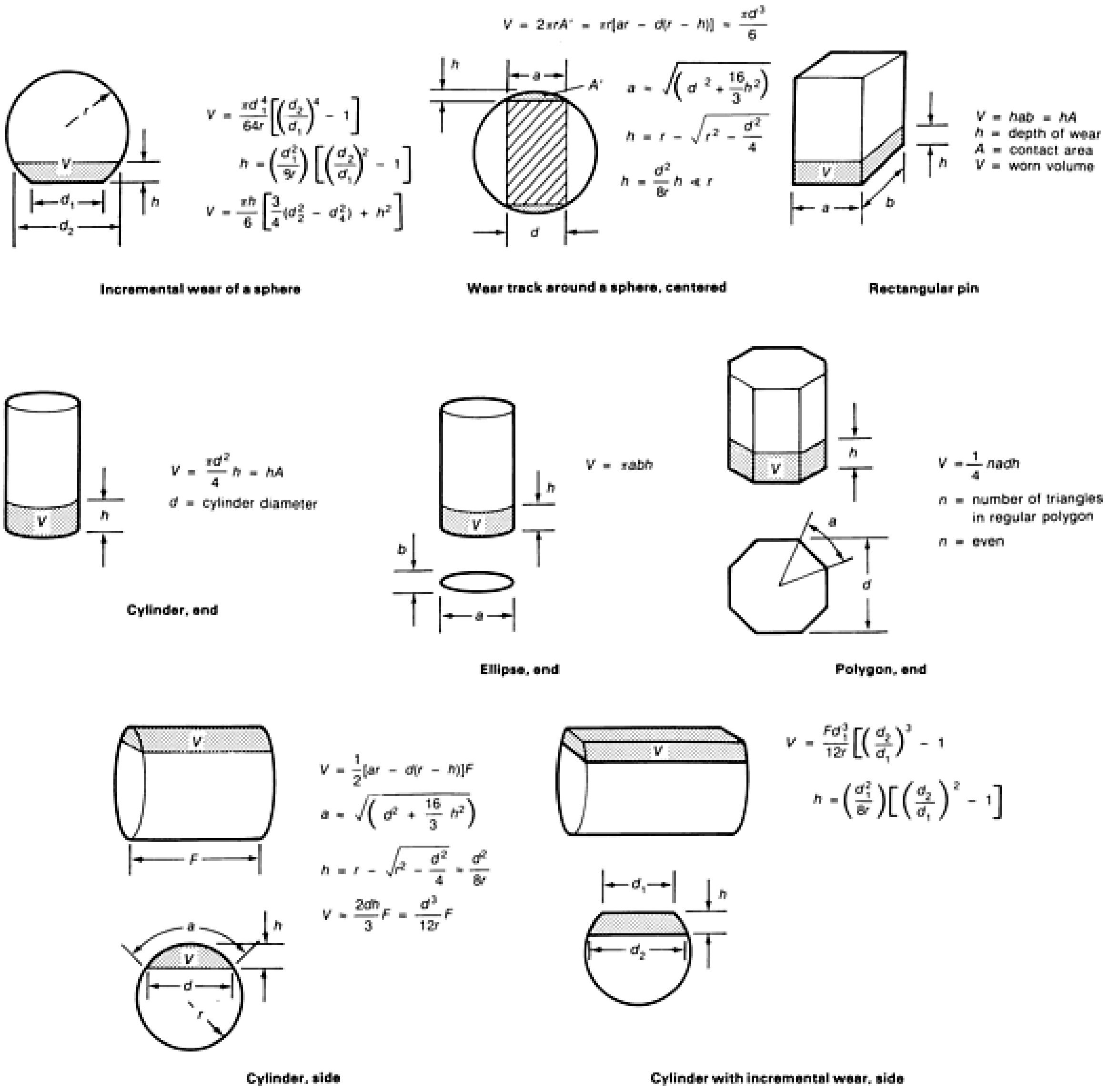


Fig. 17 Wear volume calculations of selected workpiece configurations subjected to contact with flat surfaces. Source: Ref 8

It should be noted that all of these relations assume that the opposing surface is planar and does not change its shape during the test. This is frequently not the case in wear testing. Thus, the accuracy of these equations for determining the wear volume is frequently inadequate. A more accurate approach is to determine the initial and the final shape of both specimens by surface profiling, and then to calculate the volume removed using those data.

References

1. *Wear of Materials,* Vol 1-8, American Society of Mechanical Engineers, 1977-1991
2. D. Scott, Ed., *Wear,* Vol 13, *Treatise on Materials Science and Technology,* H. Hermann, Ed., Academic Press, 1979
3. *Annual Book of Standards,* Vol 03.2, ASTM, 1990

4. M.J.H. Ruscoe, A Predictive Test for Coin Wear in Circulation, *Wear of Materials,* American Society of Mechanical Engineers, 1987, p 1-11
5. M. Matsunaga, Y. Ito, and H. Kobayashi, Wear Test of Bucket Teeth, *Wear of Materials,* American Society of Mechanical Engineers, 1979, p 336-342
6. "Wear Testing with a Pin-on-Disk Apparatus," G 99, *Annual Book of ASTM Standards,* ASTM, 1990
7. A.W. Ruff and S. Jahanmir, "Measurements of Tribological Behavior of Advanced Materials: Summary of U.S. Results on VAMAS Round-Robin No. 2," NISTIR 89-4170, National Institute of Standards and Technology, 1989
8. Wear Volume Equations, Appendix 1, *Wear Control Handbook,* American Society of Mechanical Engineers, 1980, p 449-457
9. W.A. Glaeser, Wear Measurement Techniques using Surface Replication, *Wear,* Vol 40, 1976, p 135-137
10. L.K. Ives, National Institute of Standards and Technology, private communication
11. "Ranking Resistance of Materials to Sliding Wear using Block-on-Ring Wear Test," G 77, *Annual Book of ASTM Standards,* ASTM, 1983
12. P.K. Mehrotra, Mechanisms of Wear in Ceramic Materials, *Wear of Materials,* American Society of Mechanical Engineers, 1983, p 194-201
13. E.P. Whitenton and P.J. Blau, A Comparison of Methods for Determining Wear Volumes and Surface Parameters of Spherically Tipped Sliders, *Wear,* Vol 124, 1988, p 291-309
14. P.A. Swanson, L.K. Ives, E.P. Whitenton, and M.B. Peterson, A Study of the Galling of Two Steels Using Two Test Methods, *Wear of Materials,* American Society of Mechanical Engineers, 1987, p 49-58
15. L.K. Ives, M.B. Peterson, and E.P. Whitenton, "The Mechanism, Measurement, and Influence of Properties on the Galling of Metals," NISTIR 89-4064, National Institute of Standards and Technology, 1989

Light Microscopy

W.A. Glaeser, Battelle Memorial Institute

Introduction

WEAR SCARS are sometimes complex in morphology and can indicate various wear modes by their characteristics. Wear scars will contain microscopic features like extrusions, transfer patches, pits, and fine scratches. Therefore, microscopy is essential in the study of wear processes and the identification of wear modes. Diagnosis of wear failures has parallels with fractography. There are recognizable patterns in wear scars that help to classify the wear mode, just as there are characteristic patterns for ductile-dimple fracture and intergranular fracture.

Recognition and classification of wear modes require some experience in the interpretation of the microtopography of wear surfaces. For instance, true adhesive wear, or the transfer of material from one surface to another, is often mistaken for abrasion. Initial adhesive wear can leave long striations in the surface that look like abrasive wear marks. In the abrasive wear of mining machinery, it is necessary to distinguish between high-stress abrasion and low-stress abrasion (the former being a condition where the abrasive media are crushed in the contact zone) when making a choice of materials for wear control. For example, detecting the difference between solid-particle erosion and liquid-drop erosion in a centrifugal air compressor is essential to the solution of excessive impeller damage because it means choosing between the elimination of moisture condensation in the compressor or selecting a material resistant to erosion by hard particles. It is obvious that the latter condition represents a much more difficult and expensive solution to the problem.

Low magnification light micrographs are often used for permanent records of wear failures during a failure analysis. In fact, it is highly recommended that photographs at between 5 and 10× magnification be taken before any other form of analysis is carried out. Often, in the hurry to track down a possible cause for a failure, the part is destroyed and its original condition is a matter of someone's memory.

In research, close examination and definition of the microtopography of wear scars is very useful in finding out about wear mechanisms. The investigation can include surface analysis, analysis of the morphology of wear debris, and examination of near-surface microstructure by metallographic sections and by transmission electron microscopy (TEM) of foils made from sections through wear scars. Usually, several types of microscopy must be used to define a wear pattern. These include light microscopy, scanning electron microscopy (SEM), transmission electron microscopy, and interferometry. In addition, it is essential that the basic principles of tribology are applied to the interpretation of the wear process, both in failure analysis and wear research.

Light microscopy is an essential part of the characterization of a wear scar. It should be used in the analysis to provide an initial idea of the surface structure and its relation to the function and the component of the assembly as a unit. It is essential that both mating surfaces involved in the wear process be examined. A sample of wear debris is also very helpful in making the final diagnosis.

Light microscopy provides some features for analysis not available by electron microscopy analysis. The viewer can see colors that indicate the nature of a transfer product or a lubricant decomposition film. The oxidation of a metal surface by frictional heating can be detected by the tarnish color left on the contact region. One can use a needle to probe the specimen while viewing it with a light microscope to dig out an embedded particle or to determine the friability or softness of a deposit on the surface. Polarized light can be used to determined the crystallinity of an embedded particle or a phase exposed in the surface. A magnet can be used to probe a deposit under the microscope to check for ferromagnetic material. A microindenter can be used to determine the hardness of a microscopic area exposed in a wear scar.

An oil soaked surface can be viewed to determine the distribution of a film of oil. Polymers and elastomers, which require special preparation for electron microscopy, can be easily examined with the light microscope.

Light Microscopy Equipment

The following equipment should be available for microscopy analysis of wear scars:

- Pocket magnifier (10× is a good choice)
- Stereoscopic microscope (magnification from 2.5× to 60 or 90×)
- High-quality metallographic microscope with high resolution, good depth of focus, capable of magnification up to 1000× without oil immersion, movable *x*-*y* stage, and generous space under the objective lens for bulky specimens.
- Metallograph with Nomarski attachment and high-quality photographic capability
- Ultrasonic cleaner
- Needle probes
- Surface replication kit

A complete metallographic laboratory including mounting, sectioning, polishing, and etching facilities for metallographic sections is a valuable adjunct to successful analysis of tribological surfaces. Often, it is just as important to know what is going on under the surface as well as what is happening on the surface.

Analytical Procedures

The macroscopic morphology of a worn surface should by analyzed initially. The size of the wear scar, its relation to the rest of the part, and the way the mating surface scar relates to it are all necessary information to have before the microscopic details of the surfaces are investigated. This indicates to the engineer where the microoptical investigation should be concentrated.

Example 1: Spall on the Surface of a Babbitt Metal Journal Bearing.

Consider, for example, the spall in a babbitt journal bearing surface (Fig. 1). The macrograph shows several important features in this bearing failure:

- The steel backing has been uncovered by the spall (the grooved surface in the middle of the spall)

suggesting a problem with weak bond between the babbitt and the steel shell

- A "footprint" left by a large chunk of the babbitt that came out during the spalling process can also be seen (nose that the profile of the footprint matches the shape of the spall)
- Fine cracks in the vicinity of the spall indicate areas that require closer examination, perhaps through the use of metallographic cross section. (The fine lines running up and down on the micrograph are marks caused during removal of the bearing from the assembly)

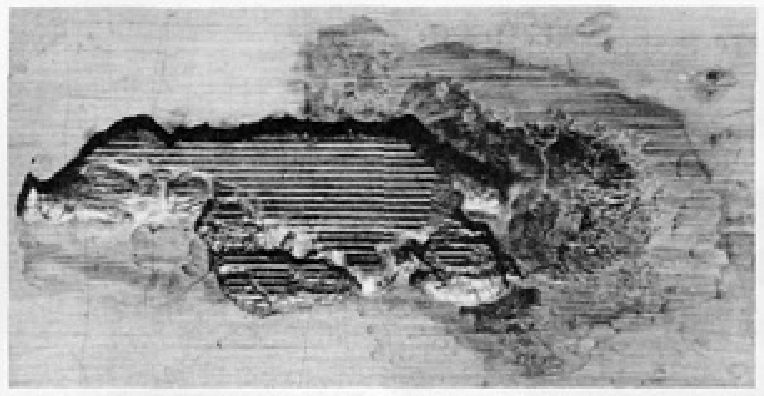


Fig. 1 Macrograph of a spall in a babbitt metal journal bearing. 8×

Example 2: Contact Fatigue Induced Spall in a Bearing Race.

Parts should be examined with a hand magnifier and with the stereoscopic microscope to obtain a general idea of the nature of the surface damage or scarring. A preliminary diagnosis of the wear mode might be made at this time. The spall mark in the bearing race shown in Fig. 2 indicates contact fatigue. The cause for the spalling is not apparent, and therefore, more detailed microscopy needs to be done in the area of the spall. In addition, the wear marks made in a bearing race can be used to detect misalignment, excessive thrust load, faulty installation, lubricant failure, or fretting.

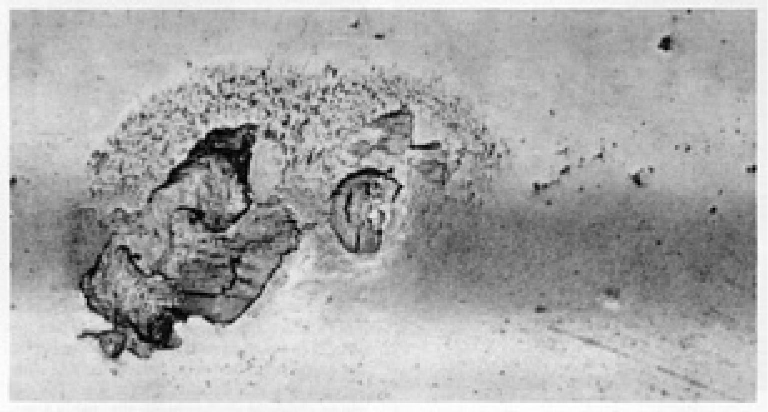


Fig. 2 Spall in a ball bearing caused by rolling contact fatigue. 10×

Example 3: Wear Scars Generated in Wire Strand Segments of Wire Rope in a Dragline Excavator.

Figure 3 shows wire segments cut from the interior strands of a wire rope from a dragline excavator. The wear scars shown on each of the three segments of wire are indicative of three contact conditions on the wire rope. The top wire contains a long scar typical or crosswire contacts where wires cross at low angles. The second wire contains cross-wire contact scars from wire contact at high angles. The third, or bottom, segment contains scars from longitudinal wire contact within the weave. All three segments indicate severe internal wear of the wire rope. The top segment shows signs

of adhesive contact, suggesting failure of the rope lubricant. The scars on the bottom segment are essentially the result of plastic deformation or bedding in of the rope assembly.

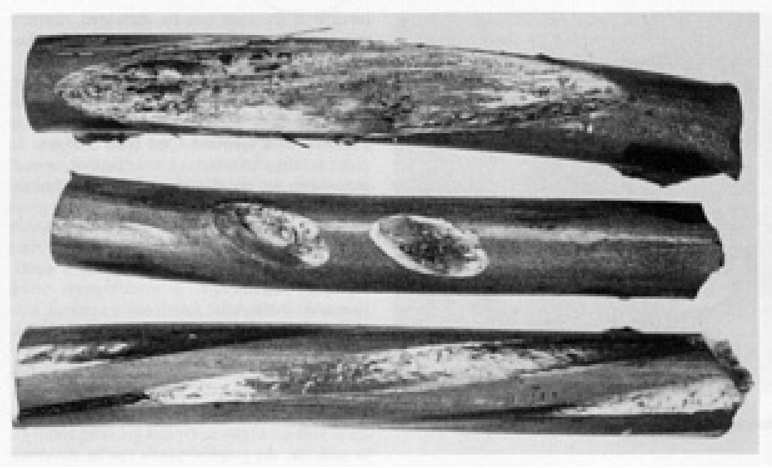


Fig. 3 Wear scars in steel wire segments cut from inner strands of a severely internally worn wire rope used in dragline operations. (Top) Long scar typical of low-angle cross-wire contact. (Middle) Scar typical of high-angle cross-wire contact. (Bottom) Longitudinal scar caused by plastic deformation of contacted wires within the weave

Example 4: Assessment of the Failure of a Ball Bearing Race Based on A Fluting Pattern Generated by Electrical Discharge and Vibration.

The fluting pattern in the ball bearing race in Fig. 4 provides an example of how a macrograph can be used for analysis of a failure. The fluting, or evenly spaced marks, in the race are the result of electrical discharge between the balls and the race surface while the bearing is experiencing a vibration mode. Using the photograph, the number of marks per inch on the surface of the race can be measured, thereby providing a mechanism for estimating the frequency of vibration.

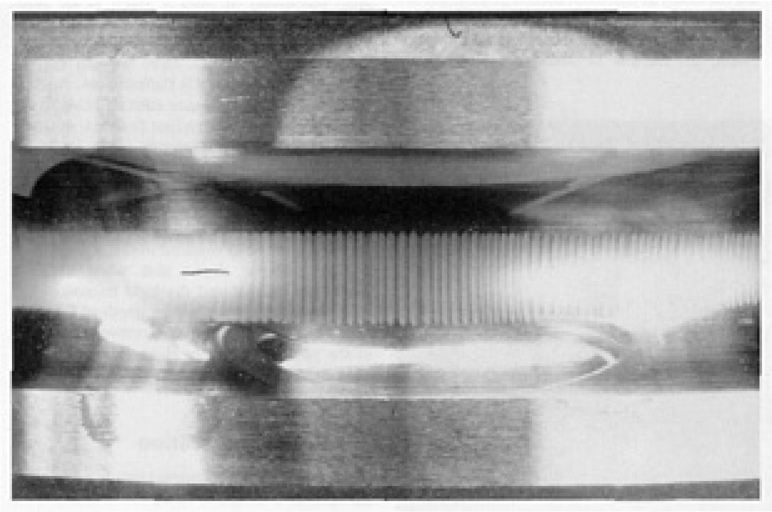


Fig. 4 Fluting pattern in a ball bearing race caused by an electrical discharge between the rolling elements and the race during mechanical vibration. 2 ×

Example 5: Application of Metallographic Microscopy, Nomarski Contrast Illumination, and SEM Techniques to Analyze Wear Patters in a Ground Cast Iron Surface.

In order to fully characterize a wear pattern, several techniques are required because one technique (metallographic microscopy, for example) gives only part of the picture. This can be illustrated in the three micrographs shown in Fig. 5. A ground cast iron surface was examined with a metallographic microscope, a metallograph using Nomarski differential interference contrast illumination (Ref 1), and a scanning electron microscope.

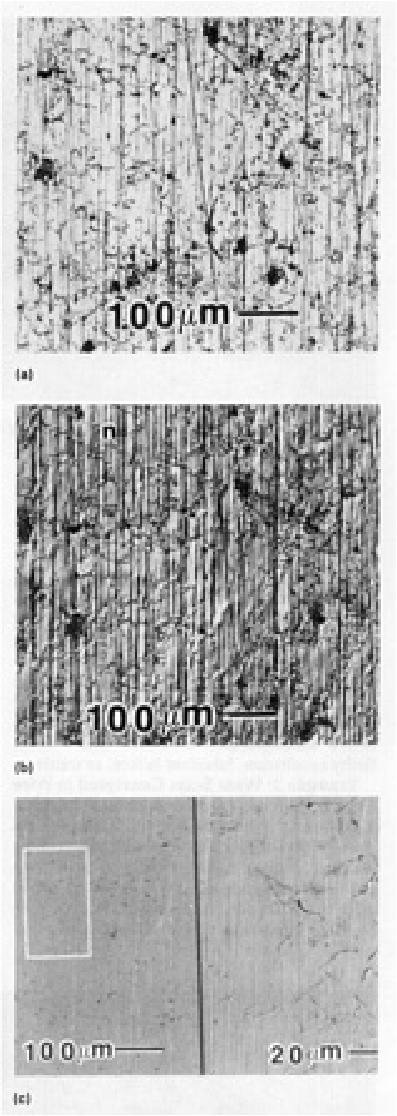


Fig. 5 Ground cast iron surface as viewed by three optical methods. (a) Metallograph. (b) Nomarski contrast illumination. (c) SEM micrographs at two magnifications

Figure 5(a), taken at 100× in a metallographic microscope, shows a fine scratch pattern, which some scratches cutting across the others at various angles (a typical grinding pattern). In addition, the graphite phase can be discerned as fine line segments at random angles. The dark spots are assumed to be shallow pits.

When the same area is viewed using Nomarski contrast, more details of the surface topography can be seen (Fig. 5b). Nomarski contrast illumination at relatively low magnification can show shallow surface relief on reflective surfaces. Figure 5(b) shows that the graphite phase is revealed by lips of metal smeared over the graphite flakes exposed end-on.

When the same surface was examined with the scanning electron microscope, some features do not show up but others become clear. Figure 5(c) is an SEM split-field photomicrograph showing two different magnifications of the same area. The field on the left is at 100× and the field on the right is of the are shown by the rectangle taken at 350× magnification. Note that the scratch pattern, so prominent in Fig. 5(b), can hardly be seen at 100× in the SEM micrograph (Fig. 5c). One does not begin to see good detail until a higher magnification is used, as the left field shows. Here we can see the scratch pattern--only it is flattened out. Nomarski differential interference contrast tends to exaggerate sharpness of surface

features, whereas scanning electron microscopy tends to flatten them (especially in the secondary electron imaging mode). Note that the SEM micrograph shows definite pits that are associated with the graphite phase. The SEM micrograph also shows lips of metal smeared over the graphite.

This example shows that when abrasion marks are present and the light microscope is used at magnifications where depth of focus allows the peaks and valleys to be revealed, details of the pattern are nicely characterized by ordinary light microscopy and Nomarski enhanced microscopy.

Specimen Preparation

Before any surface preparation is attempted, the specimen should be looked at carefully in the as-received condition. Using a hand magnifier and the stereoscopic microscope, one can look at the condition of the lubricant and detect the presence of dirt or other contaminants, corrosion, wear debris, cracks, and so on, in order to make an initial assessment of the wear mode and to determine the direction further investigation should take. Sometimes, if the observer is fortunate, the probable cause of surface damage can be determined by this type of examination. For instance, fretting corrosion of a carbon steel part can be detected by the presence of a red-colored greaselike material surrounding the wear scar. This material is the hydrated oxide of iron that has exuded from the fretting contact.

For closer examination of a wear scar using the metallographic microscope, the surface has to be cleaned. A surface contaminated with oil, grease remnants, or loose debris will mislead the observer when viewed at 500× with the microscope. It will appear that the surface is pitted, or important surface features can be completely obscured.

If the specimen is a metal covered with oil or grease, it should be cleaned in an ultrasonic bath containing a solvent. A detergent solution in water should be used in the ultrasonic bath for a polymer or for a surface with a stain on it. It is imperative that no fingerprints are left on the surface to be examined after the specimen is removed from the cleaning solution.

A failed part may have languished outdoors in inclement weather and become rusted before it is brought in for analysis. Rust can obscure important clues. It is possible to remove the rust without changing fine detail in a wear scar. The rust can be removed electrochemically, using a special electrolyte (Ref 2). The electrolyte is composed of a solution of Endox 214 (Enthrone Inc.) in cold water (240 mL, or 8 oz, Endox to 1000 mL, or 34 oz, water). A small amount of Photo-flow is added to the solution (the solution contains sodium cyanide and therefore requires caution in handling and should be opened only under a well-ventilated laboratory hood). The specimen is wired as the cathode in an electrolytic cell and should preferably be rotatable. The anode material is either carbon or platinum. An electric potential is applied at a current density of 250 mA/cm^2. The current is allowed to flow for about 1 min while the specimen is rotated. After the first minute, the specimen surface should be inspected. If not sufficiently clean, it should be given a second cleaning of 1 min duration. Final cleaning is done in an ultrasonic bath with a detergent, followed by a rinse in methanol.

Taper Sectioning

Taper sectioning can be used to study details of surface topography and the subsurface microstructure as it relates to surface features. Taper sectioning magnifies the fine surface topography and any transfer or oxide layers. Example 6 illustrates the use of taper sectioning to analyze abrasive wear on a coated steel surface.

Example 6: Application of Taper Sectioning to Determine Abrasive Wear of a Plasma Spray Coated Steel Surface.

A plasma-sprayed steel part has been sectioned (Fig. 6), and the nature of the coating bond to the substrate can be seen as well as the surface topography generated by abrasive wear. A profile of the surface roughness can be seen at the intersection of the plane of the taper section and the surface. The effect of porosity on the surface profile is revealed by the taper section. Note that the vertical magnification is about twice the horizontal magnification. Taper sectioning preserves much of the near-surface material that would not be seen in a right-angle section (normal sectioning for metallographic purposes causes rounding of the edges, unless the surface is plated before sawing and grinding). The surface can also be nickel plated before a taper section is made.

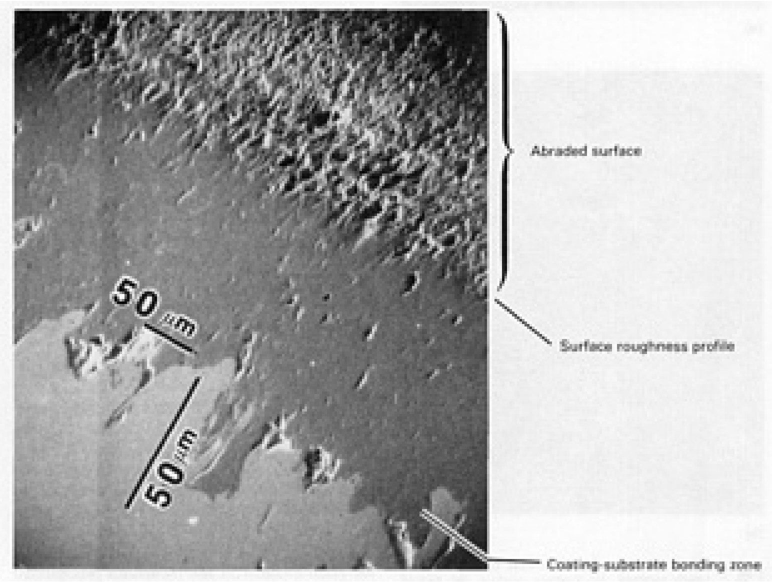


Fig. 6 Taper section of a plasma spray coated steel surface in which the surface has been subjected to abrasive wear

Example 7: Determination of Coating Thickness by Taper Sectioning for a Rectangular Block and a Cylindrical Shell.

Two types of taper section are shown schematically in Fig. 7. Figure 7(a) shows a rectangular specimen sectioned for measuring the thickness, *h*, of a coating. If the specimen is mounted for cutting so that the taper angle can be determined, the thickness of the coating revealed on the taper surface can be measured and the actual thickness determined using the equations shown below. Figure 7(b) shows a taper section made on a cylindrical specimen. Microhardness of a thin surface layer can be measured by taper sectioning because sufficient material is exposed for indenting.

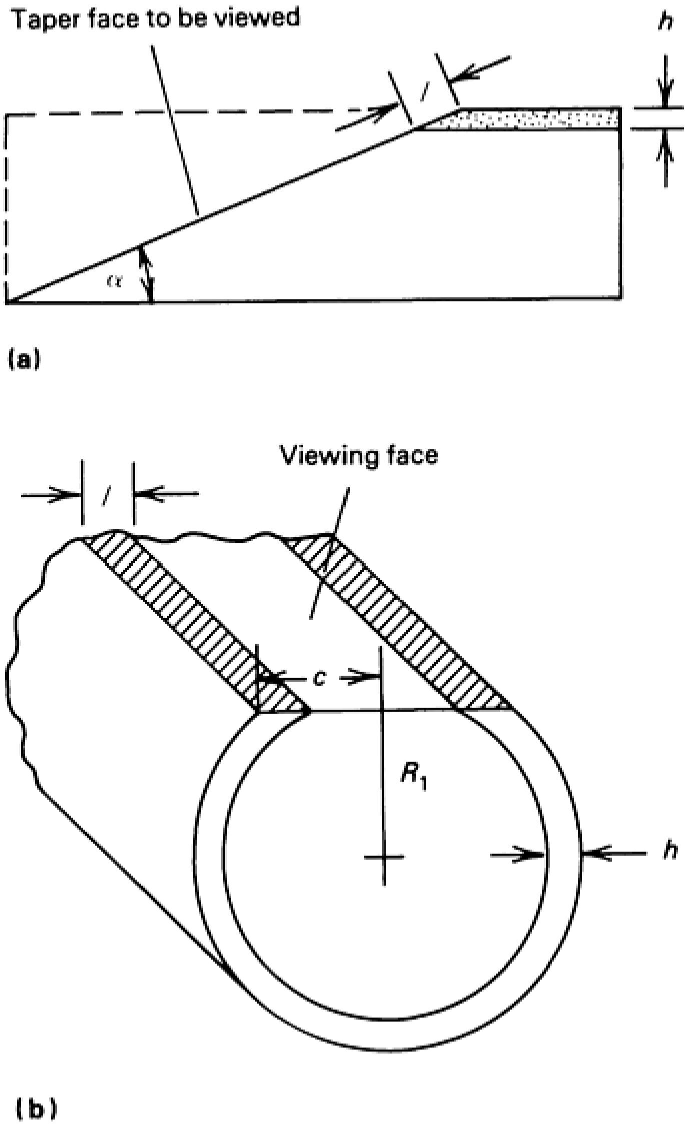


Fig. 7 Schematic showing parameters needed to determine coating thickness, *h*, of two types of taper sections. (a) Rectangular block. (b) Cylindrical

Direct geometrical measurements from a photomicrograph of a taper section on a rectangular block can be converted to real dimensions by the equation:

$$h = l \sin \alpha \qquad \textbf{(Eq 1)}$$

where h is the vertical thickness, l is the measured thickness on the photomicrograph, and α is the taper angle.

Direct geometrical measurements from a photomicrograph of a taper section on a cylinder can be converted to real dimensions by using the equation:

$$h = l \cos \left[\tan^{-1} \left(\frac{R_1}{c} \right) \right] \qquad \textbf{(Eq 2)}$$

where h is the coating thickness and dimensions l, R_1, and c are as shown in Fig. 7(b).

Surface Replication

Surface replication is a useful tool for examination of surfaces that, because of excessive size, cannot be placed directly under the microscope. It is also used for making a permanent record of a wearing surface so that the morphology of the wear scar can be recorded as a function of running time. Samples of wear scars from machinery in the field can be brought back to the laboratory for microscopy as surface replicas.

Film Replicas. Acetate films have been used for surface replication ever since electron microscopy became a practical reality. The microscopic surface features are so well recorded on the acetate film surface that high-resolution micrographs are possible at magnifications as high as 100,000×. Thus equal fidelity of surface replication for examination in the metallograph can be expected.

Film replicas of large, rough wear scars or fracture surfaces can be made with dental cement. These replicas provide a permanent record of a failure and can be used for macrographs. The area and volume of a wear scar can be determined by measurements on dental cement replicas (Fig. 8). Dental cement comes as a powder that must be mixed with a liquid and then applied to the surface as a putty and allowed to harden.

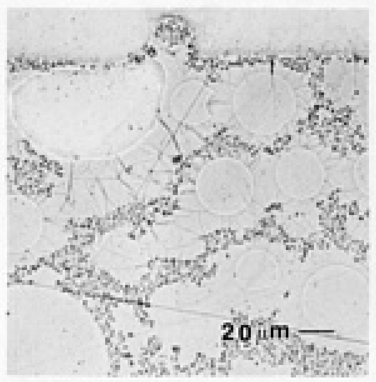


Fig. 8 Plastic-replica of a 24 μm (960 μin.) groove in a metal surface obtained using dental cement

Film replicas can be made by a number of techniques. The technique used by electron microscopists requires cutting a piece of acetate film to size, softening the film with acetone, and then pressing against the surface and holding the film in place until it hardens. The area containing the feature of concern that is recorded on the film can then be cut out and mounted on a grid for coating with gold or carbon. This process requires a certain amount of skill and is time consuming. However, if done properly, it produces the highest fidelity surface recording of any of the techniques.

Acetate Cement Films. A simpler film replica technique involves the use of an acetate cement (for example, Duco). A small drop of cement is smeared across the surface in question. This can be done by taking a drop on the finger and quickly wiping the finger over the surface to be replicated. The cement film should be thin because thick coatings can develop bubbles in them. After the cement film dries, a piece of Scotch tape is firmly pressed against the film (rubbing the tape with the eraser end of a pencil will seat it firmly against the acetate film). The Scotch tape is then pulled off with a quick snap, thus pulling off the cement film. Difficulty will sometimes be encountered in removing the replica film with tape. This is usually caused by extremely dry, clean surfaces or by interlocking of the film with surface scratches. This can usually be remedied by smearing a thin coating of oil on the surface to be replicated.

A glass slide is prepared by attaching a piece of double-stick tape to the slide. The Scotch tape with the replica on it is then attached to the double-stick tape surface so that the cement film is face up. This is gently flattened out and then viewed with the metallographic microscope. The replica surface is best viewed at 250 to 500× magnification, using the same light system as is used for viewing metallographic specimens. Although the high points are the valleys and the low points are the ridges on the plastic replica, the surface features appear normal when viewed in the microscope.

Example 8: Comparison of Surface Replica Versus SEM Micrograph of a Honed Cast Iron Gear Tooth Surface.

Figure 9(a) is a photomicrograph of a surface replication of a honed cast iron gear tooth; this replication can be compared with an SEM micrograph made of the actual gear surface shown in Fig. 9(b). Note that metal particles, which are visible as tiny white spots in the micrograph, have been picked up from the gear surface by the replica.

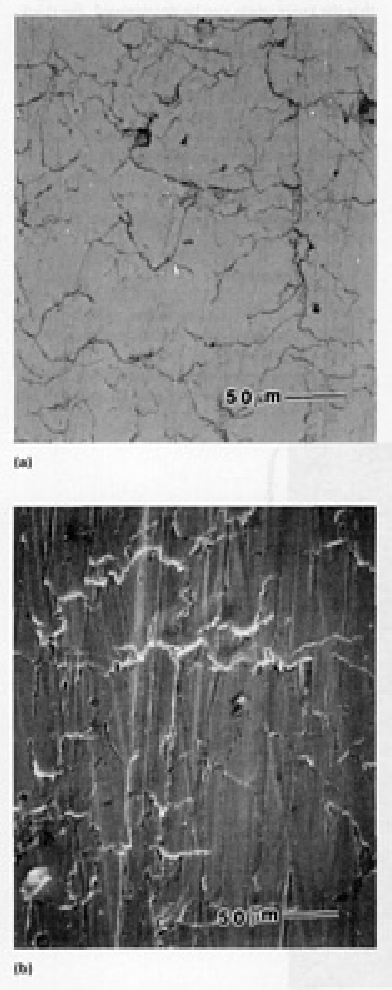


Fig. 9 Photomicrographs of honed cast iron gear tooth surface. (a) Surface replica. (b) Actual surface view obtained with scanning electron microscopy

Example 9: Application of Replicas to Remove Loose Metal Particles from the Machined Surface of a Surgical Instrument for Subsequent SEM and EDXA Analysis.

Replicas can be used to remove loose metal or other particles from a machined surface for metallographic study. This process was used in a study of the origin of metal particles in a surgical instrument (Ref 3). Metal particles picked off of the instrument surface were similar to those shown in Fig. 9(b). These particles can be located using scanning electron microscopy and identified by energy dispersive x-ray analysis (EDXA). Before the replica is examined with scanning electron microscopy, it should be carbon coated by vapor deposition to make it electrically conductive.

Metallographic Sections

Wear occurs in three dimensions. Not only is the surface involved, but the subsurface zone in the wear scar region is also significantly changed. The subsurface is subjected to very high plastic strain and microstructural changes do occur.

Subsurface voids and laminations are generated during the wear of some materials. The ferrite phase in nodular iron tends to extrude around the graphite nodules and flow to the surface. During rolling contact fatigue, carbides decompose and white etching phases develop in the maximum shear stress zone in the near-surface region. Subsurface crack patterns develop in ceramics as a result of excessive contact stresses. Because of the above phenomena, examination of the

subsurface wear scar region is usually required to completely define the wear process. Regular metallographic sectioning is used with the precaution of nickel plating the wear surface to preserve the near-surface features. Taper sectioning can also be employed when very shallow features cannot be resolved in a normal section. Subsurface features associated with the wear process are shown in Fig. 10, 11, 12.

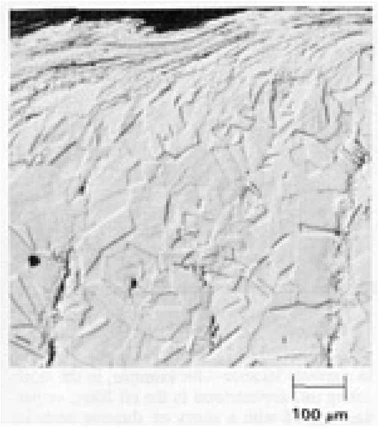


Fig. 10 Metallographic section through a wear scar that was generated by a ring-on-block wear test on a type 304 stainless steel block. Specimen was etched and the texture enhanced through the application of Nomarski contrast illumination.

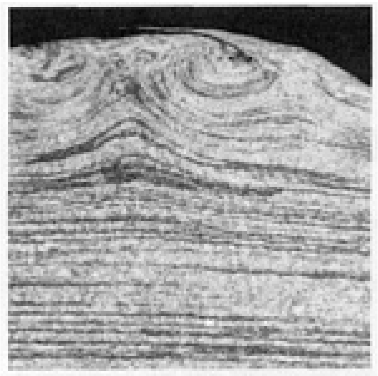


Fig. 11 Etched cross section of a hardened steel link component from a chain hoist that shows subsurface deformation caused by contact wear. 25×

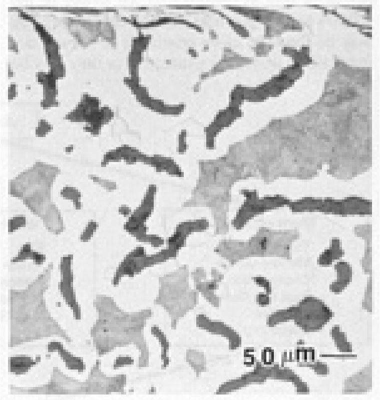


Fig. 12 Cross section of a wear scar on a nodular iron component to show deformation of the ferrite phase

(white regions) that surrounds the graphite phase (dark regions)

Example 10: Use of Metallographic Sections to Detect Subsurface Deformation in a Wear Scar Generated by Block-on-Ring Testing of a Type 304 Stainless Steel Component.

Figure 10 shows a section through a wear scar on type 304 stainless steel produced by a block-on-ring test. The section was etched and the texture enhanced by using Nomarski contrast illumination. Subsurface deformation can be seen in several stages. In the zone near the edge of the ring contact, slip lines can be seen in the metal crystals. Further back in the contact zone, much heavier deformation and flow is indicated. Some evidence of folding in of surface material into the subsurface structure can be seen. Extrusion of the flowed metal from the contact region can be seen as "fingers" of metal that break off to form debris. In the extrusion area, laminations are visible. The lamination process can be seen as the drawing out of grains into long flat plates; the grain boundaries and inclusions precipitated along the grain boundaries make up the laminations.

Example 11: Subsurface Deformation in a Chain Hoist Hardened Steel Load Link Caused by Contact Wear.

Figure 11 shows a cross section of a load link in a chain from a chain hoist. When etched, the wire forming pattern in the link can be seen. At the top of the micrograph, in the center, the deformation pattern from wear can be seen. The link surface, loaded against the mating link surface, was subjected to a rocking motion under load. The resulting subsurface deformation tended to push metal toward the center of contact to form a mound. The extrusion at the surface indicates the direction in which the link wore.

Example 12: Cross Section of a Wear Scar in a Nodular Iron Component Generated by Sliding Contact.

The third micrograph (Fig. 12) illustrates the effects of sliding contact on a nodular iron material. Note the dark phase (graphite) surrounded by a white phase (soft ferrite). Deformation and flow of the ferrite at the surface can be seen as well as cracks generated in the stretched-out graphite phase.

Wear Debris

The collection and analysis of wear debris is an important element in the determination of wear modes. The debris often contains clues to the wear mechanism and the cause of the wear. Size, shape, and color of wear debris are all indicators of the wear mode. Wear debris can be collected from the lubricant by the use of ferrography, centrifuging, and filtering.

Ferrography is useful and most successful for separating ferromagnetic debris from an oil sample (Ref 4). Ferrography separates the debris according to the size of the particles so that the distribution of debris sizes can be determined.

Filtering collects all debris, both metallic and nonmetallic, and may uncover some abrasive contaminants that might be responsible for unexpected abrasive wear. In addition, some abrasive contaminants can initiate a wear process that embeds sharp particles in a soft bearing material to cause machining of the counter surface. Filtered debris such as that shown in Fig. 13 is typical for the abrasive wear of a brittle material.

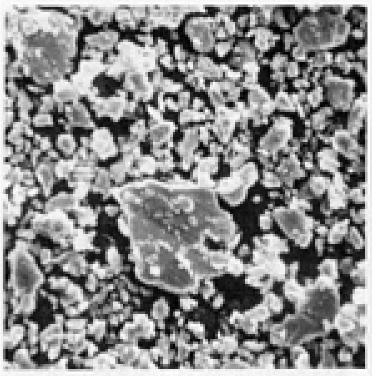


Fig. 13 Wear debris generated by the abrasive wear of a brittle material

Very fine debris can be filtered out of a lubricant after centrifuging and diluting with a solvent using a Nuclepore membrane filter. This can then be washed with a solvent and mounted on a grid for electron microscopy. A light microscope can then be used to examine the larger particles first. Wear debris from a bronze bearing test is shown in Fig. 14. The debris suggests a combined deformation and abrasive wear process. Some curled minichips that are evidence of cutting abrasion can also be identified. When this sample was viewed with a transmission electron microscope, another phase of wear debris was found: submicron size metal particles embedded in an organic gel or soap (Ref 5).

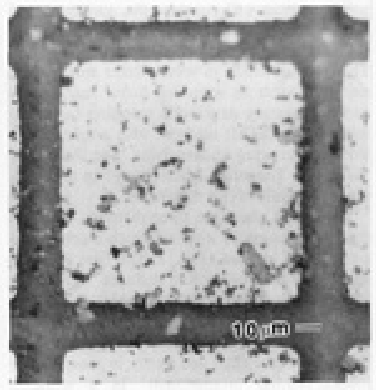


Fig. 14 Bronze wear debris generated during a lubricated block-on-ring test. The debris was obtained by centrifuging the oil through a Nuclepore membrane filter.

References

1. G.F. Vander Voort, *Metallography: Principles and Practices,* McGraw-Hill, 1984
2. P.M. Yuzawich and C.W. Hughes, An Improved Technique for Removal of Oxide Scale from Fractured Surfaces of Ferrous Materials, *Practical Metallography,* Vol 15, 1978, p 184-195
3. Battelle Memorial Institute, unpublished research
4. Wear, *Proceedings of the International Conference on Ferrography,* Vol 90 (No. 1 and 2), 1983
5. W.A. Glaeser, The Nature of Wear Debris Generated During Lubricated Wear, *ASLE Trans.,* Vol 26 (No. 4), 1983, p 517-522

Electron Microscopy

Lewis K. Ives, National Institute of Standards and Technology

Introduction

A COMPREHENSIVE UNDERSTANDING of wear processes requires a thorough knowledge of the nature and changes experienced by the contacting bodies. The first stage in the investigation of a wear problem consists of collecting all pertinent data concerning contact parameters, wear rates, friction forces, and the surrounding environment. The next step is usually the observation of the worn surfaces by means of an optical microscope. In most research laboratories, this is routinely followed by more extensive examination using a scanning electron microscope (SEM). The great depth of field and high resolution of the SEM compared with the optical microscope, and its capability for determining composition by means of associated x-ray analysis facilities, have made the SEM the premier instrument utilized in the study of worn surfaces. In fact, of the 101 papers in the most recent proceedings of the International Conference on Wear of Materials (Ref 1), 63 contained SEM micrographs, and many of these included x-ray analysis information collected in the SEM. When more detailed information about the microstructure is required, a transmission electron microscope (TEM) may be employed; three of the above-cited conference papers (Ref 2, 3, 4) contained TEM micrographs.

This article will consider the application of scanning electron microscopy and transmission electron microscopy (also abbreviated SEM and TEM) to the study of wear. Other microscopy methods, such as scanning Auger microscopy and the recently developed scanning tunneling microscopy (STM) method, also employ electrons for imaging and analysis. See the article "Scanning Tunneling Microscopy" in this Volume for details of this technique.

In general, there are three different facets in the investigation of wear to which electron microscopy is applied (Fig. 1). The first involves the study of the contacting surfaces. In a power transmission system, for example, these surfaces may include the teeth of mating gears, the ball and race components of rolling-element bearings, or the shaft and sleeve in the case of a journal-bearing arrangement. In a situation where abrasion or solid particle erosion is encountered, examination of the abrading or eroding particles as well is almost always of vital importance. Particle size, shape, composition, and microstructure often have a controlling influence on the wear process. Because changes in the surface caused by wear are usually a critical issue, the analysis will typically include a comparative examination of a similar or adjacent surface region that has not been subjected to wear. (It should be noted here that use of the term "surface" is somewhat ambiguous, referring not only to the outer boundary of the body but also to a volume of material within the body below the boundary that has been affected by tribological contact. This may include reaction films, transferred material, and compacted debris, as well as original material of the body. When specific reference is made to the region below the boundary, the term "subsurface" is often used.)

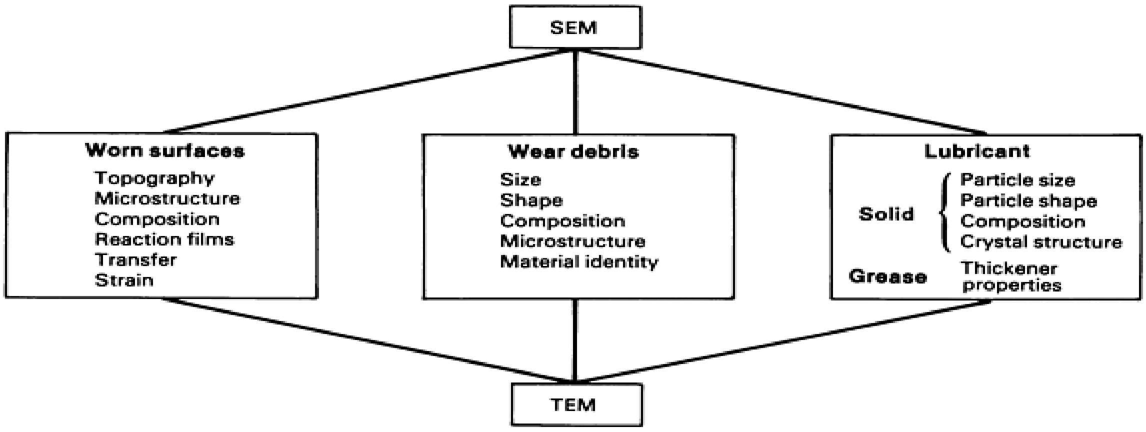


Fig. 1 Three different facets of wear to which SEM and TEM are applied

The material or "debris" that is removed from the surface constitutes the second facet of wear to which electron microscopy is applied. The shape, size, composition, and microstructure of the debris can provide important clues about the nature of the operating wear mechanisms, the conditions in the contact where wear is occurring, and the influence of the environment. Also, debris particles are not merely generated but often participate in and may even play a dominant role in determining the nature of the friction and wear processes. The debris particles may be found in close proximity to the contact from which they were generated, or they may be in a remote location--for example, in the lubricating oil, accumulated in the oil filter, or perhaps mixed with a slurry of abrasive particles that was responsible for the wear. Collection of the debris particles for examination may range from quite routine to nearly impossible, as, for example, when they are widely dispersed in an abrasive slurry medium.

In some cases a third facet, the lubricant, can be included in the microscopic analysis. In the case of solid lubricants, electron microscopy is an important analytical tool (Ref 5). Normally, electron microscopy is not applied directly to the study of oils and greases because of their high vapor pressures and rapid degradation in the electron beam. However, they can be studied using cryomicroscopy methods, where the specimen is mounted on a cold stage (Ref 6). The solid thickener component of greases has long been studied by electron microscopy (Ref 7).

Electron microscopy has played a substantial and critical role in virtually all areas of materials science. A number of text and reference books are available, covering the subject in nearly all levels of detail and thoroughness. The bibliography at the end of this article includes some of the more recent general references on SEM (Ref 8, 9, 10, 11, 12) and TEM (Ref 13, 14, 15, 16, 17, 18, 19, 20). Two shorter summary reviews on SEM (Ref 21) and TEM (Ref 22) and two references that review both SEM and TEM together (Ref 23, 24) are included for those desiring a briefer treatment than is provided in the general texts but requiring considerably more detail than is given here. Quinn has written two monographs, an early volume (Ref 25) and a more recent one (Ref 26) that deal with the application of analytical techniques, including electron microscopy, to tribology. Reference may also be made to a review by Ruff *et al.* (Ref 27) that is directly concerned with the application of SEM and TEM to the study of wear.

Two societies are devoted to the advancement of the field of electron microscopy: the Microscopy Society of America (MSA) (formerly the Electron Microscopy Society of America) and the Microbeam Analysis Society (MAS). Both groups actively promote informational and educational programs. MSA has developed and maintains a videotape library, providing information and instruction on a wide range of microscopy techniques and applications. These videotapes are available to society members for a small rental fee.

The discussion that follows is organized into several sections. First, electron-specimen interactions are briefly discussed, because this is the basis by which SEM and TEM methods are able to reveal the characteristics and properties of materials. This is followed by a short overview of TEM and SEM instruments and a summary of specimen preparation methods, particularly in connection with wear. Finally, SEM and TEM methods are considered in somewhat greater detail, with examples from actual wear investigations.

Because of the large amount and ready availability of information on electron microscopy, as well as the vastness of the subject, only a brief outline will be given here. The references should be consulted for greater detail.

Electron-Specimen Interactions

In both TEM and SEM, an electron beam is directed onto a specimen. Interactions that may occur are depicted schematically in Fig. 2. Each of these interactions is a source of information that can be exploited to reveal specimen properties. Through control of the incident electron beam and use of an appropriate detection system, unique information can be obtained regarding specimen surface topography, crystal structure, lattice defects (for example, dislocations and stacking faults), the distribution and morphology of phases, magnetic domain structure, the presence of an electrical charge, voltage variations, atomic composition, and atomic bonding. The information can be presented in the form of an image, a graph, or as tabulated data.

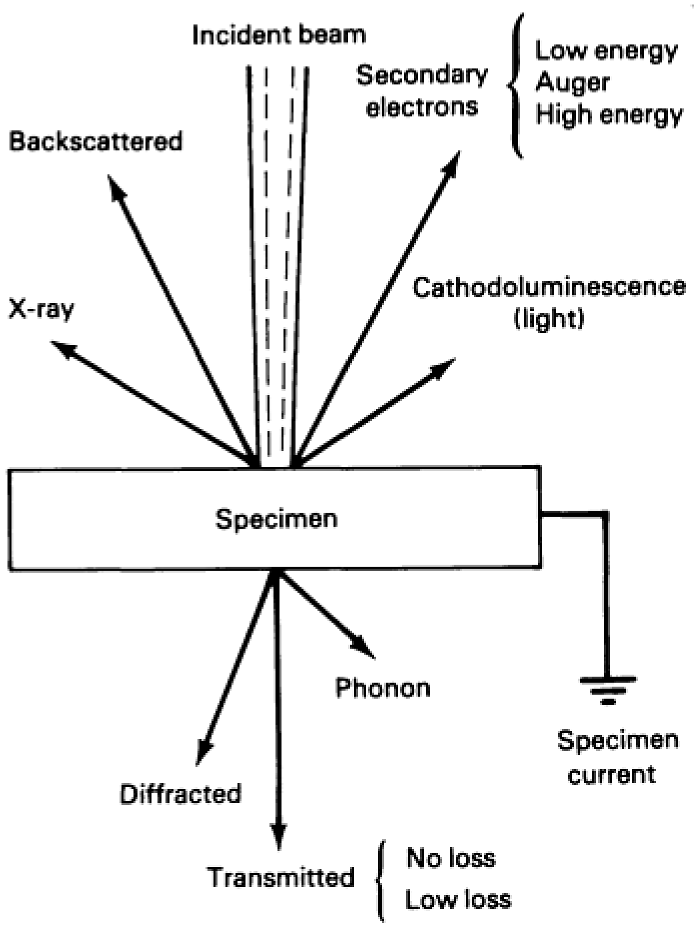


Fig. 2 Electron-specimen interaction signals

Sometimes the investigator can use this information with little knowledge of its origin. Usually, however, meaningful interpretation requires a basic understanding of the electron-specimen interaction process. A rigorous treatment of the various electron-specimen interactions generally lies in the realm of physics, particularly quantum mechanics. This approach can be greatly simplified by treating electrons and atoms as simple charged particles and applying the principles of Newtonian mechanics. Alternatively, when interference phenomena or diffraction is involved, the electrons are treated as plane waves and the principles of geometrical optics are employed. The following brief discission offers only the simplest explanations. For a detailed discussion, see the cited references.

The interaction processes in Fig. 2 can be categorized as elastic or inelastic. In an elastic interaction, the incident electron does not lose energy, but its direction is changed. In an inelastic interaction, the incident electron does lose energy and its direction is usually changed as well. Elastic scattering occurs as a result of one or more encounters with the positive Coulomb field (screened by inner-shell electrons) of atomic nuclei. Because of the comparatively large mass of the nucleus, the incident electrons may be deflected without appreciable loss of energy. The angle of deflection can reach 180°; that is, the incident electron undergoes complete reversal in direction. Elastically scattered electrons that leave the specimen contribute to the backscattered signal in Fig. 2. It should be noted that collisions with atomic nuclei at the energies available in high-voltage TEMs may result in considerable energy transfer--to the extent that atoms in the specimen are displaced, resulting in radiation damage. In that case, of course, the scattering would be inelastic.

Interaction with specimen electrons is the predominant source of inelastic scattering. The specimen electrons may acquire considerable energy in the process and thus participate in additional scattering interactions themselves. The scattered specimen electrons are referred to as secondary electrons. Those secondary electrons that leave the specimen are the source of the signal so designated in Fig. 2. One type of inelastic interaction that is especially important in connection with chemical analysis occurs when an inner-shell electron is ejected from an atom, leaving the atom in a high-energy, ionized state. For this to happen, the incident electron must have an energy that exceeds the critical absorption energy for

the electron shell of the particular atom involved. The presence of such absorption edges in the energy spectrum of inelastically scattered electrons is the basis for chemical analysis by means of electron energy loss spectroscopy (EELS).

De-excitation of the ionized atom by means of an electron transition from an outer shell to the inner vacant shell may be accompanied by the ejection of an Auger electron or the emission of a characteristic x-ray. The former serves as the basis for Auger spectroscopy and the latter for x-ray spectroscopy. Other inelastic responses are the collective excitation of conduction and valence electrons to produce plasmon waves, lattice excitations to produce phonons, and the production of electron-hole pairs in semiconductors. All of these are discussed later.

Diffraction is a direct consequence of the wave nature of the electrons and results from the interference between electrons scattered by atoms in the periodic array of a crystalline solid. Incident-beam electrons as well as previously scattered electrons are subject to diffraction.

If the specimen is sufficiently thin, electrons may leave the back surface. Incident electrons that have lost little or no energy and are not deflected during their passage are often referred to as primary-beam transmitted electrons.

The spectrum of energies for the different categories of electrons that are emitted from a surface exposed to an energetic beam of incident electrons is summarized schematically in Fig. 3. Because inelastic scattering processes may involve a range of different energies and because a given electron may be subject to several different scattering events, the spectrum is continuous--ranging from zero to the energy of the incident beam. By convention, the term "secondary electrons" is often used to refer to emitted electrons with energies less than or equal to 50 eV, and the term "backscattered electrons" is applied to electrons that leave the surface with energies greater than 50 eV, including high-energy secondary electrons, Auger electrons, low-loss electrons, and elastically scattered electrons. Many of the electrons that leave the surface have undergone multiple scattering by several different processes.

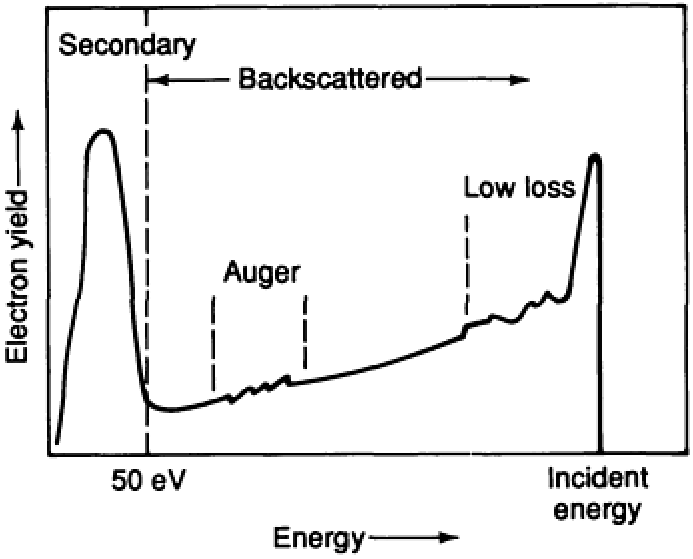


Fig. 3 Energy spectrum of emitted electrons

X-rays are produced by two processes. "Bremsstrallung" x-rays are generated by the deceleration of energetic electrons in the potential field of the specimen and have a continuous range of energies. Characteristic x-rays, as mentioned earlier, result from the relaxation of atoms ionized by the loss of an inner-shell electron and have discrete energies characteristic of the atoms in the specimen. A characteristic x-ray emitted as a result of an electron transition to a vacant K-shell site is referred to as a K-series x-ray. More specifically, $K\alpha$ x-rays are associated with the L- to K-shell transition and $K\beta$ x-rays with the M- to K-shell transition. Similarly, such series exist for other shells. Practically speaking, only the K, L, and M series give rise to x-rays that have sufficient energy to be useful in SEM- and TEM-based analysis systems.

Cathodoluminescence refers to the emission of light (infrared, visible, and ultraviolet) due to irradiation by electrons. A ubiquitous example of cathodoluminescence is the visible light emitted by the phosphor of a cathode-ray tube (CRT)

under the action of its internal electron beam. Semiconductors, various other inorganic materials and certain organic materials exhibit cathodoluminescence. Metals are only weakly cathodoluminescent. The most extensive application of the cathodoluminescent signal has been in connection with the study of semiconductors. In semiconductors, the emission of light occurs as a result of the recombination of electron-hole pairs either intrinsically between the valence and conduction band or extrinsically because of the presence of donor and acceptor sites. Discussions of cathodoluminescence can be found in Ref 8, 9, and 11.

Specimen current, I_c, originates from the absorption of electrons in the specimen. Thus, referring to Fig. 2, the specimen current is given by:

$$I_c = I_i - (I_{bs} + I_s + I_t + I_{diff}) \qquad \textbf{(Eq 1)}$$

where I_i is the incident-beam current, I_{bs} is the backscattered current, I_s is the emitted secondary electron current, I_t is the transmitted electron current, and I_{diff} is the diffracted-beam current. For a bulk specimen, I_t and I_{diff} are zero.

Finally, the electron beam gives rise to heating. This thermal signal is not used directly, but has been exploited in connection with thermal wave imaging, whereby an acoustic signal resulting from the expansion caused by heating is detected.

An important concept in connection with the beam-specimen interaction process is that of the interaction volume. This is the volume of material penetrated by the incident beam and includes elastically and inelastically scattered electrons. For specimens of low atomic number, the volume has a teardrop shape, as shown schematically in Fig. 4. The volume becomes more hemispherical in shape for materials of higher atomic number. The relative depths from which the various signals may escape to the surface are indicated in Fig. 4.

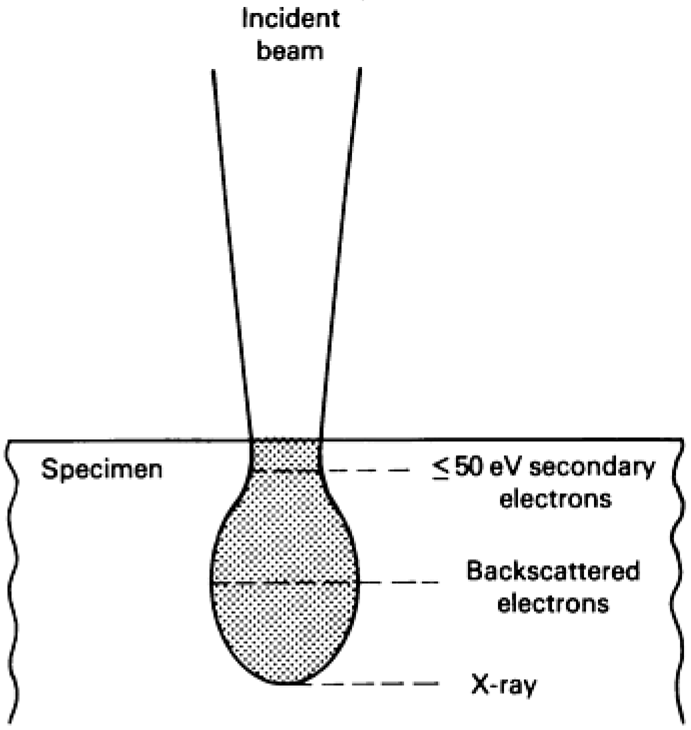


Fig. 4 Electron-specimen interaction volume

SEM and TEM Instruments

The first transmission electron microscopes were designed solely to extract information from transmitted and diffracted electrons beams, and the first scanning electron microscopes utilized secondary and/or backscattered electrons. Modern

TEMs and SEMs continue to be optimized for these signals; however, they also make extensive use of several other signals shown in Fig. 2. For instance, both microscope make use of the emitted x-ray spectrum to obtain information on composition, greatly enhancing their analytical capabilities.

Schematic drawings illustrating the basic configurations of SEM and TEM instruments are shown in Fig. 5(a) and (b), respectively. Both instruments employ an illumination system consisting of an electron gun and subsequently electromagnetic lenses to project an appropriately focused and/or collimated beam of electrons onto the specimen. In the TEM, additional lenses below the specimen are provided for image formation and magnification. Suitably located apertures are employed to defined the beam, along with coils for beam alignment, astigmatism correction, and beam scanning. A vacuum environment is required for both microscopes to limit scattering by gas molecules.

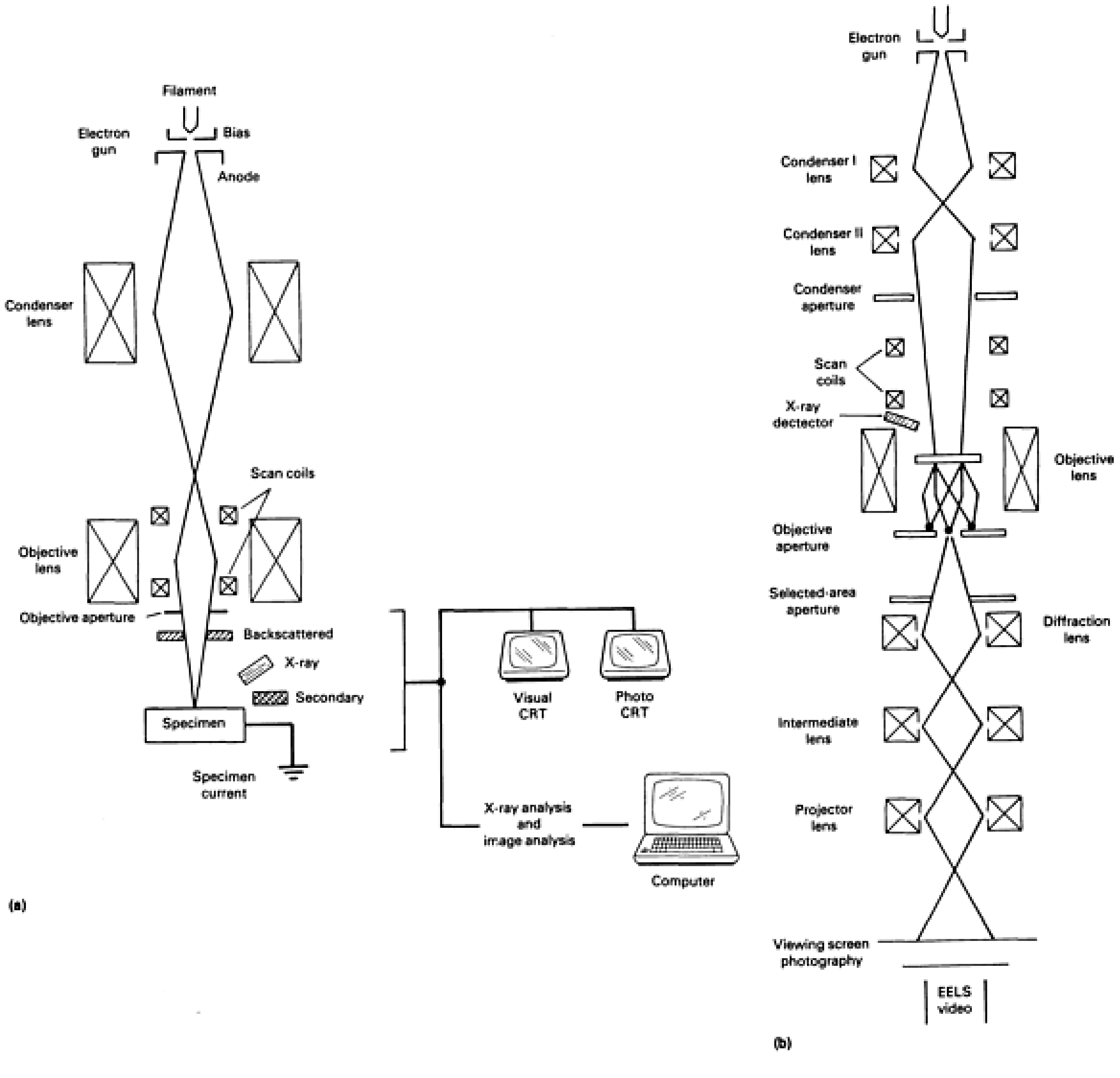


Fig. 5 Basic configuration of (a) SEM and (b) TEM

The most common source of electrons in both instruments is a heated tungsten filament, although single-crystal lanthanum hexaboride (LaB_6) emitters and tungsten field emission tips are gaining in popularity. These two sources are substantially brighter than the tungsten filament, as much as 10^2 times for the LaB_6 source and 10^4 for the tungsten field emission source. In general, they provide improved resolution and analytical capabilities. Their operational life is also greater, but cost is higher and a more sophisticated vacuum system is required.

The electrons are accelerated through a potential that is established by an applied voltage between the filament and anode. SEM accelerating voltages typically range from about 100 V to 30 kV. Higher accelerating voltages are not ordinarily

used with the SEM, because this instrument is applied primarily to the study of surface features of bulk specimens. In general, as will be discussed later, an accelerating voltage is selected that best suits the application at hand.

Transmission electron microscopes are available in three different accelerating voltage ranges. The most commonly used instruments operate at a maximum of 120 kV, but allow the selection of voltages as low as 20 kV. With so-called intermediate-voltage microscopes, the maximum voltage is typically 200 to 400 kV. High-voltage instruments are capable of operating at 10^6 V and higher.

In general, a higher accelerating voltage permits penetration of thicker specimens and provides improved resolution. However, the gain in going from the intermediate-voltage range to the high-voltage range is relatively small for all but the most specialized applications and is achieved at a very substantial increase in cost. Intermediate-voltage instruments allow routine observation of the atomic structure of all classes of crystalline materials. This, together with increased penetration, improved EELS capabilities, and the fact that specially constructed laboratory facilities (necessary for high-voltage instruments) are not required, has led to an increase in the popularity of intermediate-voltage instruments.

In the SEM, the specimen is normally located below the final lens in the illumination system. For improved resolution, however, some instruments provide a second position within the final lens. The electron beam is focussed to a small spot and scanned serially over the specimen to form a rectangular raster. Secondary electrons (or one of the other sources in Fig. 2) are collected to provide a signal that is amplified and used to modulate the intensity of the electron beam in a CRT, The CRT beam is scanned in synchronism with the electron beam incident on the specimen, resulting in an observable image. The magnification of the image is determined by the ratio of the distance scanned on the specimen to the corresponding distance displayed on the CRT, the latter generally being kept constant. Thus, when the scanned area on the specimen is small, the magnification is high, and when the scanned area is large, the magnification is low.

Resolution in the SEM is determined to a first approximation by the diameter of the beam incident on the specimen. In practice, the diameter of the beam is controlled by the type of filament, the excitation of the condenser lenses, the final aperture size, and the position of the specimen with respect to the final lens. The latter distance is referred to as the working distance. A critical factor limiting the useful probe size is the available beam current. Increasing the excitation of the condenser lenses or reducing the size of the final aperture results in a small electron probe diameter; however, the probe current is also reduced. This in turn reduces the strength of the signal available for amplification, so that the usable minimum probe size is limited by the capacity of the signal amplifier, that is, the signal-to-noise ratio of the amplifier. With a heated tungsten filament, the probe diameter is limited to about 5 nm; with LaB_6 and field emission sources, diameters of about 3 nm and 1 nm, respectively, are achievable.

In actual operation, the nature of the specimen and the signal source (secondary electrons, x-rays, and so on) usually play a limiting role in determining resolution. If the features of interest result in only a small difference in the signal, then an increased probe current and correspondingly larger probe size are required, thereby reducing resolution.

The TEM in its conventional mode of operation differs significantly from the SEM. The specimen is illuminated by a relatively broad, nearly parallel, stationary beam of electrons. Transmitted and diffracted electrons that have lost little or no energy and that do not deviate too far from the optical axis are focused by the objective lens. Subsequent lenses provide additional magnification and allow observation of either the image or a diffraction pattern. The image and diffraction pattern can be viewed directly on a fluorescent screen, photographed, or displayed by means of a TV system.

As an additional enhancement, TEMs have also been adapted for operation in a scanning mode, similar to the SEM. For this purpose, the illuminating beam is focused to a small probe. Detectors are included to sense transmitted and diffracted electrons, as well as secondary and backscattered electrons as in the SEM. A transmission electron microscope designed to function either in the conventional stationary illumination mode or in the scanning mode is commonly referred to as a scanning transmission electron microscope (STEM). The acronym CTEM is often used to refer to a conventional transmission electron microscope without STEM capability, or simply to operation in the conventional mode. Transmission electron microscopes are available that have been designed and optimized to operate only in the scanning mode; such instruments are usually referred to as "dedicated" STEMs. Finally, a TEM equipped with x-ray detectors, and perhaps with electron energy loss spectrometers as well, is frequently referred to as an analytical electron microscope, or AEM.

Originally, SEM and TEM instruments were essentially analog in operation. X-ray analysis systems designed as separate attachments to SEMs and later to TEMs incorporated digital technology. As development has progressed, these systems have become capable not only of processing x-ray data but also of controlling the position of the beam on the specimen.

Thus, composition maps can be acquired and stored in computer memory. With suitable programming, the stored data can be used, for example, to provide information on the proportion of each compositionally different phase in the sample or to count particles having a particular composition. In essence, the x-ray analysis system has also become an image analysis system. Accommodating other signals in the system, such as secondary and backscattered electron signals, is a fairly easy step. Thus, image analysis can be done directly with these signals, eliminating the need for indirect analysis of photographs and separate image analysis equipment.

Recent TEM and SEM instrument designs rely heavily on digital technology and computer control, incorporating keyboards and computer CRT screens. The operator interacts with a software program rather than directly manipulating an array of knobs, buttons, and switches. The image can be digitized and sent directly to computer memory or stored in a nonvolatile memory medium, such as a hard disk. As a result, image analysis and x-ray analysis capabilities are directly incorporated in the instrument.

Finally, it should be mentioned that the SEM is not merely a passive instrument for examining wear-related specimens; it may also incorporate a wear testing device for *in situ* observation of wear processes. A number of important results have been obtained in this way (Ref 28, 29, 30, 31, 32, 33).

Specimen Preparation

Scanning Electron Microscopy

An important advantage of the SEM is that a specimen can be examined with little need for special preparation beyond that required for the optical microscope. This does not imply that the condition and nature of the specimen do not have a significant bearing on the quality of the image and the information obtained. Indeed, revealing the desired information may require considerable effort with respect to sectioning, polishing, and etching.

For a specimen to be suitable for examination, it must be free of volatile matter that might interfere with the attainment of an operational vacuum level or result in the contamination of apertures and other components, thus degrading the performance of the instrument. A contaminant film that hides the surface features of interest is certainly unacceptable. Even a thin deposit of oil may lead to the rapid appearance of a dark film over the area scanned by the electron beam. This is especially noticeable when focusing is carried out at a higher magnification, leaving a telltale dark square in the lower magnification field. Solvent cleaning or low-temperature vacuum bakeout for porous specimens may be required. The surface should also be free of extraneous dust particles, which may charge or otherwise detract from the image. Specimen cleaning is discussed in Ref 8 and 12.

Before adopting a particular cleaning method, careful consideration should be given to the effect of the procedure on surface films and attached wear debris. Wear debris and triboreaction films almost always yield important information regarding wear processes. If necessary, loose films and debris may be removed for separate analysis, using methods discussed later, freeing the underlying surface for examination.

A second requirement is that the specimen must have adequate electrical conductivity to allow electrons to flow to ground or to the specimen current amplifier connector without charging. For metal and some semiconducting specimens, the inherent conductivity is sufficient, and it is necessary only that the specimen make good electrical contact with the mounting device that attaches to the microscope stage. Specimens that are poor conductors or insulators must be coated with a conductive layer, usually a heavy metal such as gold or a gold-palladium alloy.

Alternatively, if compositional or crystal structure studies are to be made, interference may be minimized by a carbon coating, assuming that carbon itself is not one of the constituents in the specimen of interest. The coating should not obscure the details of the microstructure and topography being examined. This becomes especially important in high-resolution studies, where care in the selection of coating material and coating method is of critical importance. Materials and methods for coating are discussed in several general references (Ref 8, 9, 10, 12).

In addition to the direct observation of worn surfaces in the SEM, a replica of the surface may be prepared for examination. This approach is required when the component of interest is too large to be accommodated by the SEM or cannot be sectioned, or when it is not desirable or feasible to remove the component from its system. One method for preparing a replica is to employ the cellulose acetate tape used to prepare TEM replicas. A piece of the tape is moistened with acetone and pressed against the component surface. After a suitable drying time, the tape is stripped from the surface and coated for examination in the SEM. Rough surfaces may require use of a different replicating material (Ref 12, 20). It

should be noted that the replica method is limited in its ability to provide an accurate representation of the surface. This is especially true in connection with cracks and holes, neither of which may be easily recognized using a replica. Also, a lip of folded-over material can often be identified by direct surface observation in the SEM, but usually not with a replica.

Although much can be learned by direct examination of the worn surface, the subsurface is also an important source of information. Oxide layers, films, compacted debris, cracks, deformed layers, and transformed regions may be poorly revealed or not revealed at all when the outer worn surface is studied. Examination of the subsurface requires the preparation of a section through the specimen. It is common practice to prepare a cross section perpendicular to the surface. Moreover, because relative motion between the contacting bodies usually occurs along a specific direction on the surface, it is best to prepare two sections, one parallel and the other perpendicular to the direction of motion.

In general, more information is obtained from the parallel section. The majority of material flow usually occurs in the direction of sliding, and the parallel section is best suited to reveal its pattern. If grain boundaries or other suitable microstructural features are present, it may be possible to measure the strain as a function of depth by the amount of bending or microstructural distortion (Ref 34, 35). Tensile cracks oriented with their plane roughly normal to the direction of sliding are also best observed in the parallel section.

Preparation of cross sections requires considerable care and skill. Soft materials and specimens with fragile or poorly adherent films are probably the most difficult candidates. Figure 6 is an example of a section through the wear track on a copper surface. The specimen was electroplated with a layer of copper before sectioning parallel to the direction of sliding. The deformed region below the track is clearly visible. Dislocation cells and an annealing twin are displayed as a result of channeling contrast (discussed below).

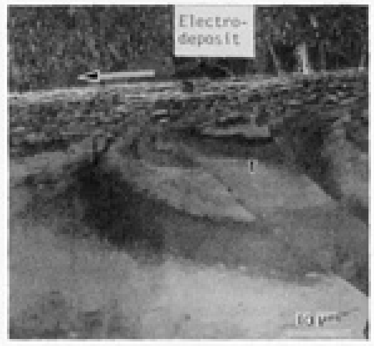


Fig. 6 Cross section through wear track on a copper surface parallel to sliding direction. T, annealing twin, and arrow indicate direction of motion with respect to counterface. Source: Ref 27

A taper section may also be prepared through the worn surface (Ref 36). This method makes it possible to obtain what is in effect a magnified view of linear and planar features on the surface, such as scratches and films, respectively. The length of the scratch or width of the film in the taper plane, together with the known taper angle, allows the scratch depth or film thickness to be determined. Note that scratches or ridges must not be parallel to the line of intersection of the taper plane and the surface and preferably should be perpendicular to that line. Furthermore, for an accurate determination, the feature of interest should be uniform in thickness or depth over the length of the taper plane.

There are a number of methods for collecting and preparing debris specimens for examination in the SEM. When oil and other volatile materials are not present, loose debris remaining on the worn surface can be studied directly be placing the component in the SEM. If the component is too large or cannot be removed from its system, the debris must be removed by stripping it away with adhesive tape or by brushing it onto the tape. Double-sided adhesive tape facilities attachment to the specimen mount. Although conductive tape may be used, it is usually necessary to apply a coating for conductivity.

The surface replica methods described above may also be used to remove debris for examination. When the debris is present in oil, a small amount may be placed in a suitable solvent, such as hexane. The oil can be eliminated by centrifuging and replacing the contaminated solvent with fresh solvent. Ultimately, a drop of debris-containing solvent is placed on a specimen mount for examination in the SEM. An alternative approach involves washing the solvent-oil-debris mixture through a porous membrane filter, which is then coated for examination in the SEM. Finally, it should also be

mentioned that debris distributed on ferrography slides (see the article "Lubricant Analysis" in this Volume) can be examined in the SEM (Ref 37).

Transmission Electron Microscopy

In contrast to specimen preparation for the SEM, specimen preparation for TEM examination almost always involves considerable effort. Because specimen preparation plays such an important role, several volumes have been written that are devoted entirely to the subject (Ref 38, 39, 40, 41).

The discussion that follows will first consider the preparation of specimens from bulk materials, including both thin sections and surface replicas. Methods used to prepare specimens from debris particles will then be reviewed.

Specimens from Worn Surfaces. The main challenge in preparing TEM specimens from bulk materials is to obtain a section that contains a region approximately 100 μm^2 or more in area with a thickness of 5 to 500 nm. Moreover, the surface of the thinned area should be smooth and free of contamination, and the internal structure should not be altered by the preparation process. For the study of defect structures, a thickness of approximately 200 nm is suitable, although thicker specimens may be acceptable for materials of low atomic number and thinner specimens would be favored for materials of high atomic number. Operation at high accelerating voltages will extend the maximum usable thickness. For high-resolution studies, the specimen should be no thicker than 40 nm; the optimum value depends on the material and the nature of the study to be performed. Because most preparation methods result in a wedge-shaped section, achieving the desired thickness is not that difficult; at some location within the wedge, a suitable thickness can usually be found.

The only distinction between TEM specimen preparation methods for tribological studies and TEM specimen preparation methods in general concerns the strong emphasis placed by the former on surface and near-surface material. However, tribology is not the only field where interest is focused on the surface. The study of corrosion, of semiconductor devices, and of surface treatment processes, such as nitriding, ion implantation, and so on, have promoted the development of methods for preparing specimens from the surface region.

A thin section may be taken parallel to the worn surface, or perhaps a sequence of several sections prepared in order to study the structure as a function of depth. If material immediately at the surface is to be studied, a soluble coating may be necessary to protect the surface during preparation. The parallel section approach to examining the immediate surface is feasible only if the worn surface is quite smooth. Otherwise, a cross section normal to the surface must be prepared. As discussed in connection with SEM specimen preparation, the orientation of the section may be chosen either parallel to a characteristic direction, such as the sliding direction, or perpendicular to that direction. An electrodeposited layer (for metal specimens) or a thick coating of cement (for example, epoxy) is usually applied to the surface for protection during sectioning and thinning. A technique commonly employed for semiconductor devices is to cement together a pair of specimens face to face (or a stack of several specimens, in the case of thin semiconductors) (Ref 42). Not only is protection provided, but having more than one specimen also improves the chances for success.

A general scheme that is often followed in preparing TEM specimens from the bulk is illustrated in Fig. 7. It is assumed that the specimen is a cross section and that the worn surface has been protected by a thick coating layer (>1.5 mm). First, a thin section is cut, with care being taken that the section is thick enough that damage from the cutting process does not extend into the region of actual study. For hardened steels and ceramics, a thickness of 01.1 to 0.2 mm may be acceptable. For soft, annealed metals, 1 mm or larger may be required. A low-speed diamond saw or, for conductive materials, an electric discharge machine (EDM) is frequently used for cutting.

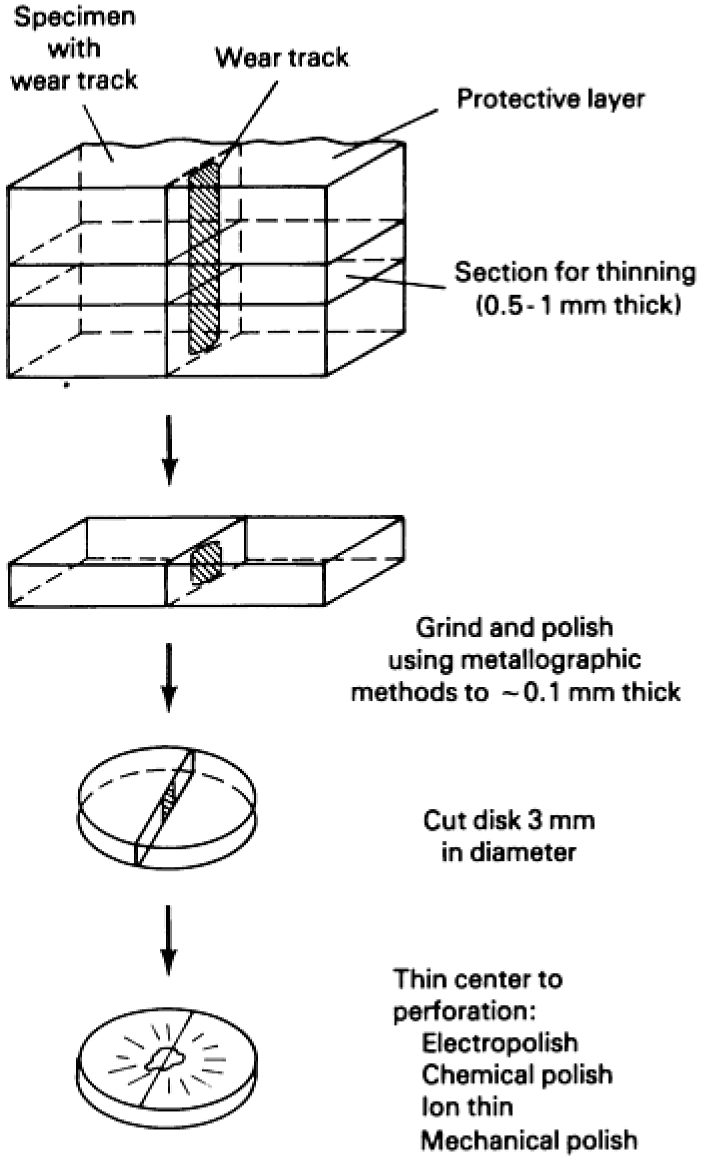


Fig. 7 Schematic showing steps required to prepare TEM cross section specimen

Then the thickness of the section must be reduced to about 0.1 mm. Conventional abrasive grinding, lapping, and polishing methods are usually employed (Ref 43). One or more disks 3 mm in diameter are cut from the thin section. Because this is a cross section, the line of intersection demarcating the worn surface is placed at the center of the disk. The disk may be cut with a hollow or tubular-shaped tool having an inside diameter of about 3 mm. One of several different cutting methods may be used, depending on the material--for example, core drilling with a tool impregnated with diamond grit, abrasive slurry core drilling, ultrasonic abrasive impact machining, or EDM.

The final step is to reduce the thickness of the center region of the disk until it becomes thin enough to be electron transparent. In practice, thinning is usually continued until perforation occurs. When successful, some of the area around the perforation will be thin enough for electron transmission. Thinning may be accomplished by electropolishing, chemical polishing, ion beam milling, possibly mechanical polishing, and combinations of these processes. Alternative methods and approaches for preparing specimens, as well as additional details, can be found in Ref 37, 38, 39, 40, 41, 42.

Surface replicas have already been discussed in connection with the preparation of specimens for SEM study. As was mentioned, a technique that is often used involves pressing a thin section of cellulose acetate film moistened with acetone onto the surface. After a short time is allowed for drying, the film is stripped from the surface. For TEM study, the acetate replica is coated (shadowed) at oblique incidence with a heavy metal such as palladium-gold or chromium. The shadow

produced by the heavy metal absorbs or scatters electrons, enabling the topography to be revealed by transmitted electrons.

After shadowing, a uniform layer of carbon, 10 to 20 nm thick, is deposited--in effect generating a carbon replica of the acetate film surface. Deposition of the shadowing metal and the carbon support film is usually done in vacuum evaporator. The shadowed and carbon-coated acetate film is cut into pieces 3 mm square. One of the squares may then be placed on a TEM support grid and the cellulose acetate film dissolved with acetone--leaving the shadowed carbon replica, suitable for examination in the TEM. This describes only briefly one of many different methods (Ref 20, 37), for preparing replicas. The same cautionary note made with respect to accurate surface representation in reference to SEM replicas is also true for TEM replicas.

Debris Specimens. Wear debris particles that are thin enough to be electron transparent can be studied directly in the TEM. In this case, preparation consists of collection, dispersion, and mounting the particles on a suitable support film. Specimen grids covered with a 10 to 20 nm thick film of carbon or silicon monoxide are often used for this purpose. If the particles are in the form of a dry powder, they can be brushed or blown such that some fall onto the support film.

Alternatively, the particles may be dispersed in a volatile solvent and deposited as a small drop or sprayed onto the support film. Particles in-oil or grease can be extracted by solvent washing through an appropriate membrane filter or by repeated centrifuging in a solvent and decanting until the particles are free of contaminants, as discussed in connection with SEM specimen preparation. When a membrane filter is employed, it is treated like a replica; that is, the filter surface is coated with a layer of carbon to support the particles, and the filter material is dissolved away.

When the wear debris particles are too thick to be electron transparent, they are usually embedded in mounting material (epoxy, for example) to form a composite, which is then sectioned and thinned like a bulk specimen (Ref 44), or they may be incorporated in a thick plating which is subsequently thinned. Additional information on particle specimen preparation can be found in texts dealing with particle analysis in general (Ref 45, for example).

Imaging and Analysis in the SEM

The most important signals that are employed for analysis in the SEM and the information that each provides are summarized in Table 1. Each signal requires an appropriate detector (except for specimen current, where the specimen itself is the detector) and an amplifier. In the case of x-rays for quantitative studies, a complete computer-based analysis system in necessary.

Table 1 SEM signals

Signal	Information	Special requirements
Secondary electron	Topography; some crystallography; some composition	None
Backscattered electron	Composition; topography; crystallography; magnetic domains	None
Specimen current	Similar to secondary and backscattered electrons	None
X-ray	Composition	Smooth surface for quantitative analysis
Cathodoluminescent	Composition	Used for materials that exhibit cathodoluminescence
Thermal wave	Subsurface defects	Smooth surface

Secondary Electron Signal

The main application of SEM is the investigation of surface topography, and the low-energy secondary electron signal ($\leqslant$50 eV) is the primary source of this information. Secondary electron emission is strongly influenced by surface orientation and varies approximately as the secant of the angle of incidence of the electron beam. An element of surface that is inclined to the beam appears brighter than one that is normal to the electron beam. Enhanced brightness is seen at sharp edges, small particles, and fine-scale roughness because of the larger area from which secondary electrons can escape. With increased penetration at higher accelerating voltages, the area also increases, so that more detailed and sharper images of fine surface features are more often obtained at lower accelerating voltages ($\sim$5 to 10 kV) than at higher voltages.

The observed contrast is also influenced by the position of the secondary electron detector, which is usually located to one side and at about the same level as the specimen. More electrons are received by the detector from an element with its surface inclined toward the detector than from an element tilted away from the detector. The visual effect is to make the image appear as though the specimen were illuminated by a source located at the detector.

Although secondary electrons are created throughout the interaction volume (Fig. 4), because of their low energies only secondary electrons originating close to the surface are able to escape and contribute to the image. The maximum escape depth ranges from about 1 to 10 nm for high- and low-density materials, respectively. Thus, the specimen surface immediately under the incident beam is the source of directly generated secondary electrons. If this were the only source of secondary electrons, the image resolution would be closely determined by the beam or spot size. However, secondary electrons are also generated by backscattered electrons as they leave the surface or strike the SEM polepiece and specimen chamber walls. Backscattered electrons have a large range and may exit the surface some distance from the location of the incident beam. This effectively increases the source size and decreases resolution.

In general, the fraction of secondary electrons emitted is relatively insensitive to the atomic number, *Z,* although some compounds do have a significant effect on emission (Ref 8). However, the number of backscattered electrons generated is quite sensitive to *Z* (discussed below). Thus, secondary electron emission is indirectly, affected by *Z* through its influence on backscattered electron emission. For this season, phases with different average atomic numbers can be distinguished in secondary electron images. Similarly, the emission of secondary electrons is not directly sensitive to crystallographic orientation, but, because backscattered electron emission is influenced, differences in orientation produce contrast in secondary electron images. A notable example is the variation in contrast exhibited by different grains in the image of a carefully polished polycrystalline sample; these grains exhibit channelling contrast. This effect can be seen in Fig. 6, where an annealing twin is visible and dislocation cells with only slight differences in orientation can be distinguished. Moreover, grains in the electrodeposited layer of copper can be seen.

Further examples of images obtained in the secondary electron imaging mode are shown in Fig. 8. The specimen is the worn sealing face of a diesel engine valve. The valve was titled at a large angle to the incident electron beam in order to display the lip of material that resulted from plastic flow. It is only because of the tremendous depth of field associated with the small angular aperture of the SEM objective lens, assisted by the dynamic focusing capability with which most SEM instruments are equipped, that such an image can be obtained. (Dynamic focus refers to the programmed change in focus as a function of raster position as the beam is scanned across the specimen.) In addition to the images in this article, secondary electron SEM images of worn surfaces can be found elsewhere in this Volume. Note especially those in the article "Surface Damage."

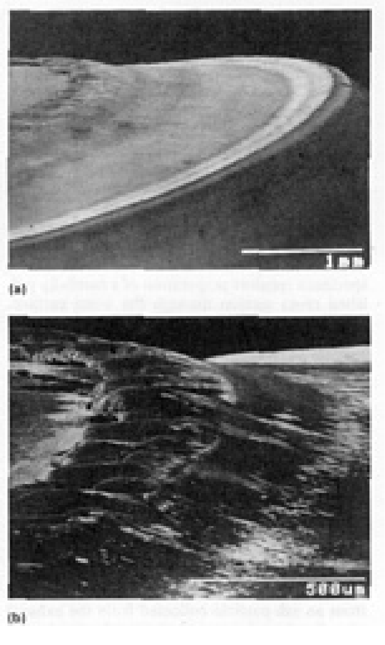


Fig. 8 Secondary electron image of the sealing face of a diesel engine valve. (a) Low magnification. (b) Higher magnification of flowed lip. Note dark contrast at carbonaceous deposit.

Other sources of contrast in addition to those discussed above are electric and magnetic fields. Both can strongly influence the number of secondary electrons collected. This permits the imaging of magnetic domains and is the basis for voltage contrast in semiconducting devices (Ref 9).

Finally, the ability to carry out stereomicroscopy using secondary electron images can be of considerable value in the examination of surface topography. Stereopairs are produced by photographing the surface at two different angles of tilt, usually at a separation of 5° to 10°. A stereoviewer can be used to observe differences in height visually, or the difference in distance between corresponding pairs of points in the two images can be measured and the height difference obtained quantitatively by simple geometry utilizing the known tilt angle and magnification.

Backscattered Electron Images

Backscattered electrons are those electrons that leaves the specimen with energies greater than 50 eV (Fig. 3) and have a component of direction opposite to that of the incident beam. This includes inelastically scattered electrons and, at the high-energy limit, primary electrons that have undergone elastic scattering with almost no loss in energy. The majority of backscattered electrons have energies from 0.5 to 0.9 of the incident beam energy.

As mentioned above, the efficiency with which backscattered electrons are generated is strongly dependent on the atomic number of the scattering atoms. This dependence is depicted schematically in Fig. 9, which shows that the scattering efficiency for light elements is less than that for heavy elements. The distribution of backscattered electrons is also a function of the beam orientation with respect to the specimen surface. Figures 10(a) and 10(b) illustrate the effect of surface orientation on distribution for normal and oblique incidence. Because of these dependencies, the backscattered electron signal carries both topographic and compositional information.

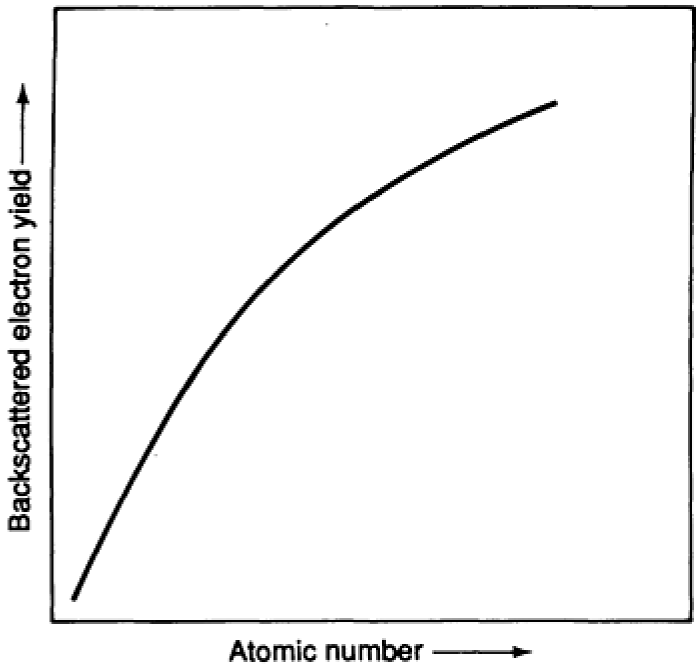


Fig. 9 Backscattered electron yield dependence on atomic number

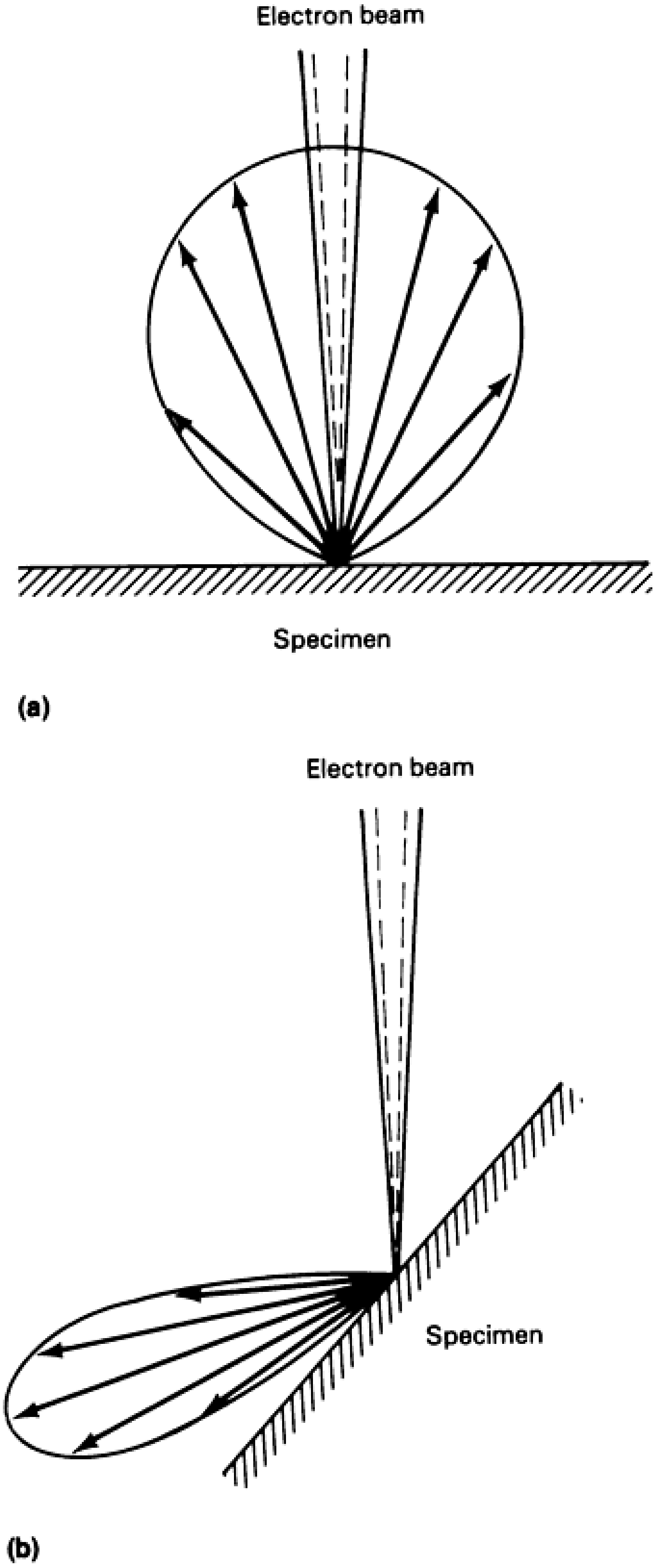


Fig. 10 Angular distribution of backscattered electrons. (a) Incident beam normal to surface. (b) Incident beam inclined to surface

The backscattered electron detector is often designed to detect electrons from several (usually four) separate locations around the specimen. The signals from each of these different quadrants may be individually selected, and added or subtracted. This feature allows the selective emphasis of atomic number contrast or topographic contrast. When the

electrons from all directions are collected and summed, the contrast is less sensitive to topography, and atomic number differences are emphasized. If the surface is very smooth, extremely small differences in composition can be detected, and the contrast is a sensitive means of mapping variations in average atomic number. Alternatively, by selecting signals from different quadrants, or adding and subtracting signals, topographic contrast can be enhanced, while atomic number contrast is suppressed. Figure 11(a) is an example of a backscattered electron image emphasizing atomic number contrast. The light features are silver transferred on a copper surface when a silver pin was slid on a copper flat. Some topographic contrast is evident along with the atomic number contrast. For comparison, the secondary electron image is shown in Fig. 11(b). In the latter image the contrast is primarily topographic.

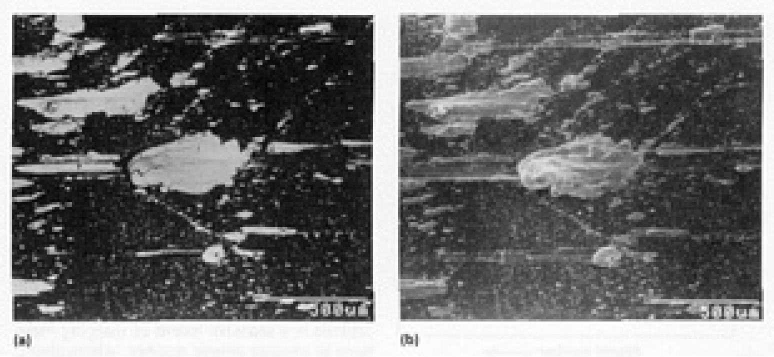


Fig. 11 Transferred patches of silver on a copper specimen surface. (a) Backscattered electron image in which silver (light contrast) with a higher atomic number yields more backscattered electrons than copper. (b) Secondary electron image of the same area

The propagation of electron waves in crystalline materials is strongly orientation dependent. This gives rise to an effect known as electron channeling (Ref 14, 15, 16, 17, 18, 19, 20, 21, 22), which relates to the enhanced flow of electrons along crystallographic planes. There is a corresponding effect on the intensity of emitted backscattered electrons. As mentioned earlier, this gives rise to the variation in contrast that distinguishes grains of different orientation at the surface of a carefully polished polycrystalline specimen.

By rocking the electron beam about a stationary point on the specimen, this same effect can be utilized to produce an electron channeling pattern (ECP). The pattern consists of light and dark bands corresponding to the crystalline planes of the material at the location of the beam. SEMs equipped to obtain ECPs are capable of generating crystal orientation lattice parameter information from areas as small as 1 μm in diameter. The quality of the ECP is influenced in a systematic way by the amount of strain in the material, and this fact has been exploited as a means of measuring strain (Ref 9). Ruff (Ref 46) has employed this method to determine the strain in the vicinity of wear tracks. If the specimen surface is heavily strained or is covered by a thick film, an ECP will not be formed.

X-ray Microanalysis and Mapping

Utilization of emitted x-rays to determine specimen composition is an extremely important capability of the SEM. The description of the method as being one of microanalysis is justified by the small volume from which information is obtained and the sensitivity of the technique. For a smooth specimen surface, the x-ray source volume is represented by a region on the order of 1 μm in diameter, about the size of the interaction volume in Fig. 4. The size can be increased by the fact that x-rays generated within the interaction volume may fluoresce additional x-rays outside the volume before reaching the surface. For fluorescence to occur, the exiting x-ray must exceed the ionization energy of the shell to which the fluoresced x-ray belongs.

Two types of x-ray spectrometers are employed. The wavelength-dispersive spectrometer (WDS) separates the emitted x-rays according to wavelength. X-rays that enter the spectrometer are Bragg diffracted by an appropriate single crystal, and the intensity is measured by means of a proportional counter. By appropriately rotating and moving the diffracting crystal and the counter, the spectrum of emitted x-rays can be scanned in serial fashion. To cover the complete spectral range, however, may require more than one crystal. The minimum concentration of an element that can be detected can be as low as 0.01% under favorable conditions.

The second type of x-ray spectrometer separates the emitted x-rays according to energy and is thus referred to as an energy-dispersive spectrometer (EDS). A semiconductor diode, most often silicon into which lithium has been diffused (called a lithium-drifted silicon detector), is used to measure the x-rays. Some application is also made of a germanium base because of its greater sensitivity to low-energy x-rays. In either case, electron-hole pairs produced by an absorbed x-ray photon result in a charge pulse proportional to the energy of the photon. The pulses are amplified, shaped, and sorted according to energy by means of the succeeding components of the analysis system. The number of pulses, or counts, is a direct indication of the emitted x-ray intensity at the given energy. The net result is a spectrum of x-ray intensity that can be displayed as a function of energy. In contrast to the WDS system, the EDS system detects x-rays of all energies as they are delivered to the detector--that is, effectively in parallel, rather than serially--resulting in an acquisition process that is much more rapid. On the other hand, EDS resolution of nearby x-ray lines (~150 eV) is (~20 eV).

The EDS detector crystal must be maintained at approximately liquid nitrogen temperature in a high vacuum. Protection from the external environment is provided by a window. The window is usually constructed from a thin film of beryllium, but polymer and diamond films are being increasingly used. Absorption in the window decreases the sensitivity of the detector to low-energy x-rays. Thus, a detector equipped with a beryllium window is limited to the detection of elements with atomic numbers equal to or greater than that of sodium (Z = 11). Carbon (Z = 6) and even boron (Z = 5) can be detected when polymer and diamond windows are used.

Windowless detectors are also available. A vacuum-tight cover is closed to prevent exposure of the detector crystal to atmosphere or poor vacuum conditions. The operating microscope environment must of course be relatively free of condensable vapors. Even in the absence of a window, absorption of low-energy x-rays takes place in the thin (~20 nm) gold layer that must be deposited on the detector surface for conductivity and in an inactive (dead) layer of silicon present on the detector crystal. This absorption, coupled with the low inherent yield of characteristic x-rays by light elements, means that the sensitivity for light elements is much less than for heavy elements.

EDS systems are simpler and less expensive than WDS systems and are the predominant type of spectrometer utilized with SEMs. In more sophisticated installations, however, both types of spectrometers may be employed with the same instrument.

Both WDS and EDS systems use computers for control of data acquisition, display, and processing. Processing involves the identification of elements represented by observed characteristic peaks and the determination of their concentrations according to the size of the peaks. Identification is relatively straightforward for large, well-separated peaks. Small peaks only slightly above background, or which overlap for different elements, make identification more difficult. In some investigations, only the identification of the elements present is sought. In other cases, quantitative concentration values are required.

In general, the theory and associated computations for determining concentrations are quite complex (Ref 8, 47). Accurate results require that the specimen surface be smooth and that the analyzed volume of the specimen be homogeneous throughout. A relatively rough wear surface covered with a film of unknown thickness does not satisfy this criterion. A more acceptable specimen requires preparation of a carefully polished cross section through the won surface. The limited spatial resolution (~1 μm^3 volume) may still restrict accuracy, especially in the presence of a transferred layer or reaction film 100 nm or less in thickness. This may be resolved by applying TEM methods (see the section "Imaging and Analysis in the TEM" in this article).

The analysis itself may be semiquantitative (without reference standards) or quantitative (with standards). In the latter case, prior collection of spectra from pure specimens of each element present or from alloys of known concentrations is required. Fortunately, the computations are done quickly and efficiently by the system software, and the process is relatively transparent to the operator. An x-ray spectrum obtained from an ash particle collected from the exhaust of a diesel engine operated on pulverized coal-fuel is shown in Fig. 12 (Ref 48). Due to the small size and complicated shape of the particle, element identification only was carried out in this case.

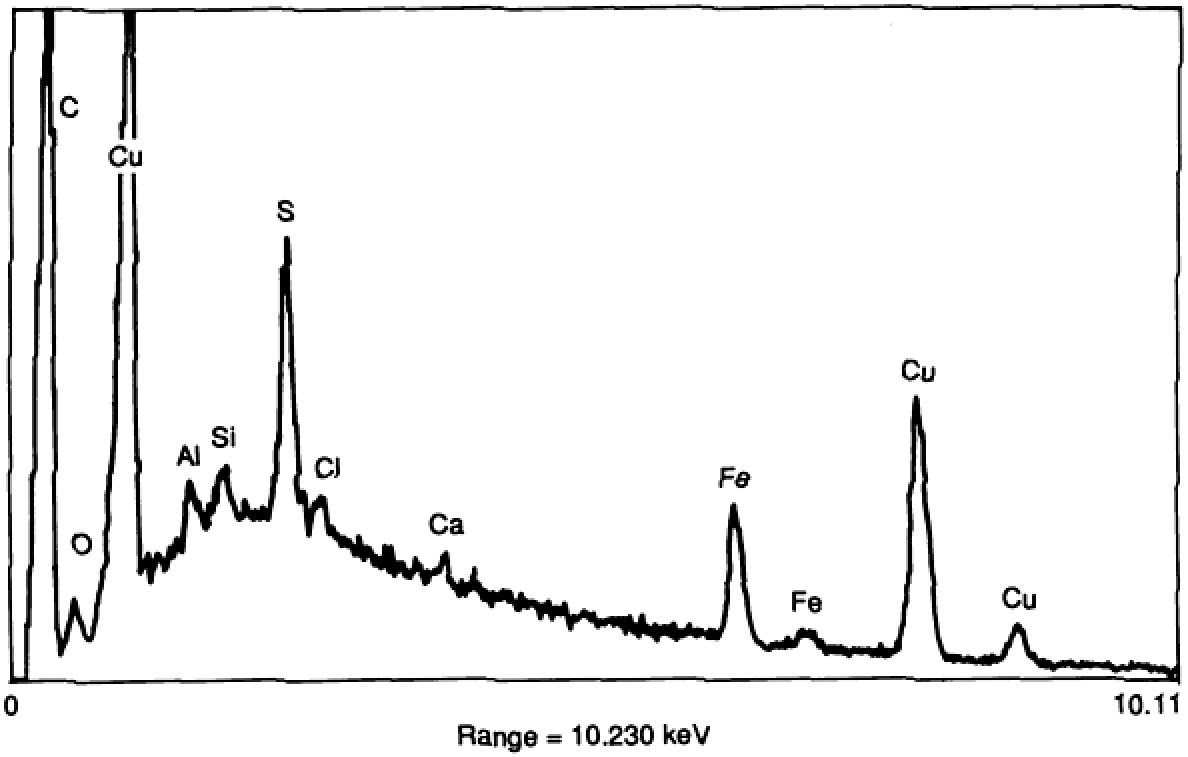


Fig. 12 X-ray spectrum from coal-fueled diesel engine exhaust particle. Copper mount is source of copper peaks. Source: Ref 48

In addition to determining the local composition of a specimen, the x-ray signal can generate a map showing the distribution of elements over a scanned area of the specimen surface. The procedure involves selecting a region of window in the x-ray spectrum that includes the elemental peak of interest and using the signal to modulate the intensity of the SEM CRT as the electron beam is scanned across the specimen. Because each x-ray photon is recorded as a pulse, the resulting image consists of a pattern of bright dots, and is usually referred to as a dot map. When the x-ray analysis system is equipped to control the position of the electron beam of the SEM (digital beam control), the map can be fully quantified at each picture element and stored directly in the analyzer computer. An example of a dot map is shown in Fig. 13(a). The specimen (the same as that shown in Fig. 11) was copper with silver transferred during sliding. The bright dots correspond to silver. A secondary electron image from the same area is shown in Fig. 13(b).

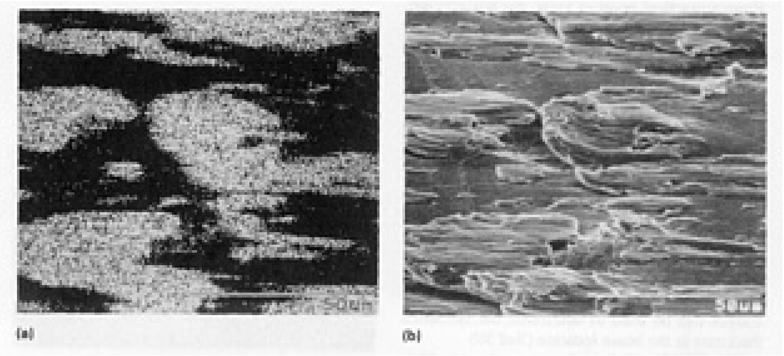


Fig. 13 Transferred patches of silver on a copper specimen surface. (a) X-ray dot map of silver. (b) Secondary electron image of same area

Other Imaging Modes

Specimen current images bear a complementary relationship to the secondary and backscattered signals (Eq 1) and can be exploited to reveal topographic, compositional, crystallographic, and magnetic information. At the extremely small currents available, especially when the probe size is made small for higher-resolution studies, the limited bandwidth of the required high-impedance, high-gain, direct-current amplifier limits observation and recording to very slow scan rates. In practice, relatively little use is made of the specimen current imaging mode.

Many minerals, semiconductors, and organic compounds exhibit relatively strong cathodoluminescence signals. Cathodoluminescence is sensitive not only to the presence of such materials but also to their properties. For example, it can be exploited to study defects in semiconductors. Relatively little use appears to have been made of cathodoluminescence in tribology studies.

The last imaging mode that will be discussed is thermal wave imaging. This mode requires a modification to the SEM that allows the electron beam to be interrupted or "chopped" at a high frequency. Heating of the specimen by the electron beam occurs at a corresponding high frequency resulting in a thermal wave. Thermal expansion causes an associated acoustic wave that can be detected by a transducer attached to the specimen. The signal is influenced by the thermal properties and microstructure of the specimen. Thus, different phases and subsurface defects such as cracks and voids can be detected. An example of an application of thermal wave imaging to wear can be found in the work of Blau and Olson (Ref 49).

Imaging and Analysis in the TEM

Basic Imaging and Diffraction Modes

Electrons incident on sufficiently thin regions of the specimen may be transmitted. The fraction of transmitted electrons depends on the accelerating voltage and on the characteristics of the specimen, including thickness, composition, density, and crystallographic orientation. Transmitted electrons that have not been scattered, elastically scattered electrons (especially diffracted electrons), and inelastically scattered electrons that have lost only a small amount of energy may be utilized to form an image. Focusing and magnification are accomplished by a series of electromagnetic lenses (Fig. 5b). The lens design is such that electrons scattered by more than a few degrees from the optical axis will not be brought to focus. The mode of operation is determined by the current that is used to energize each lens and the selected apertures. In modern instruments, the lens currents are preprogrammed according to the chosen imaging or diffraction mode. Adjustments required of the operator are mainly those associated with focusing, illumination, and magnification. A simple ray diagram construction is commonly used to represent the path of the electrons and the function of the various lenses and apertures in the different modes of operation (Ref 19) (see Fig. 14).

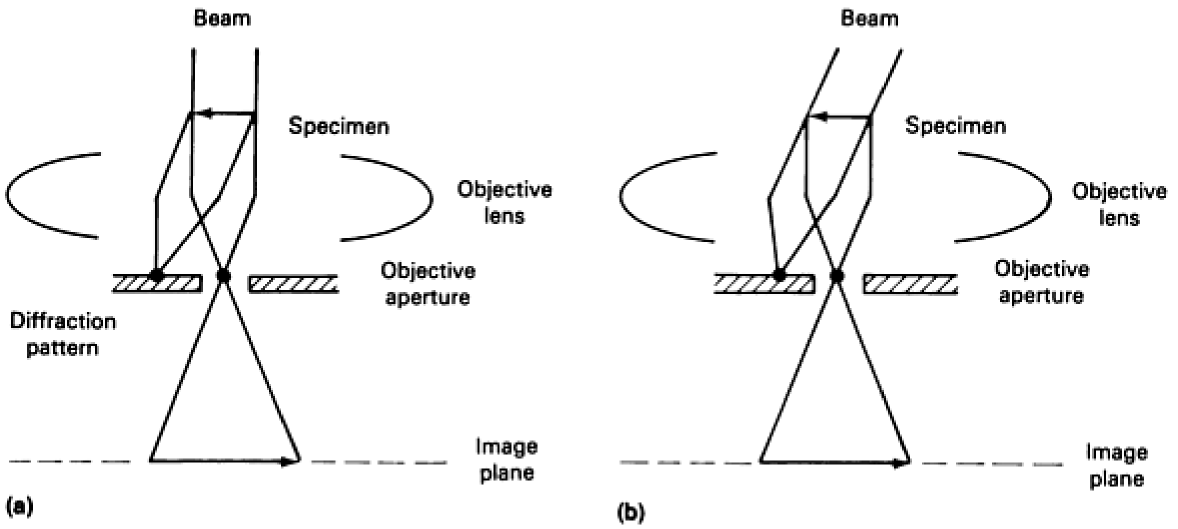


Fig. 14 TEM ray diagram illustrating (a) bright-field and (b) dark-field imaging modes

As in the optical microscope, the most critical lens in the TEM is the objective lens, which produces a magnified and focused image of the specimen. The design and quality of this lens determines the resolution of the instrument. The focal length of the objective lens is very small, 1 to 3 mm, and the specimen is immersed in the magnetic field of the lens. In achieving the desired lens characteristics, rather severe restrictions are imposed on the size of the specimen holder and the specimen that it must accommodate. The holder is usually designed to accept a specimen disk with a diameter of 3 mm, although 2.3 mm is sometimes employed. The maximum thickness of the disk is usually limited to approximately 0.5 mm or less. Precise translation, tilting, and rotation are necessary capabilities of the specimen holder and its associated stage.

Lenses following the objective lens are used to control magnification and to switch between the imaging and diffraction modes. Both the image and the diffraction pattern are observable on a fluorescent screen below the final lens. Photographic film exposed directly to the beam can be used to provide a permanent record of the image. Alternatively, a TV imaging system can be used. A combination of TV imaging and image intensification facilitates the achievement of precise instrument alignment and focusing required in high-resolution work. Also, through connection to a video recorder, the time sequence of a dynamic experiment can be recorded.

Before TEM image contrast is discussed, electron diffraction should be considered. Electron diffraction is not only a means of determining crystal structure and orientation but is also closely associated with the process of image formation.

Electron Diffraction

A transmission diffraction pattern of the illuminated specimen region is always present in the back focal plane of the objective lens. When this plane rather than the image plane is brought into focus by the first intermediate or diffraction lens (the terminology varies with manufacturer), the diffraction pattern rather than the image is displayed. Operationally, switching between imaging and diffraction is accomplished by little more than the press of a button. There are several different operating modes for obtaining a diffraction pattern; these are briefly discussed below. Greater detail can be found in Ref 19. The references should also be consulted for full details on diffraction theory and such closely related and very important topics as crystallography and crystal structure determination.

Under conventional imaging conditions--that is, when the specimen is illuminated by a large-diameter, nearly parallel beam of electrons--the area of the specimen from which the diffraction pattern is obtained is determined by an aperture introduced into the image plane of the first intermediate or diffraction lens. This mode of diffraction is referred to as selected-area diffraction (SAD). SAD patterns are shown in Fig. 15(a) and 15(b). The selected area in Fig. 15(a) included only a single grain, while many grains with different orientations are included in Fig. 15(b), giving rise to spots in rings. For these patterns, the angle 2θ between the incident beam (the center spot) and each of the diffracted beams is given by Bragg's equation:

$$2d \sin \theta = n\lambda \qquad \textbf{(Eq 2)}$$

where d is the lattice spacing, λ is the electron wavelength, and n is the order of diffraction.

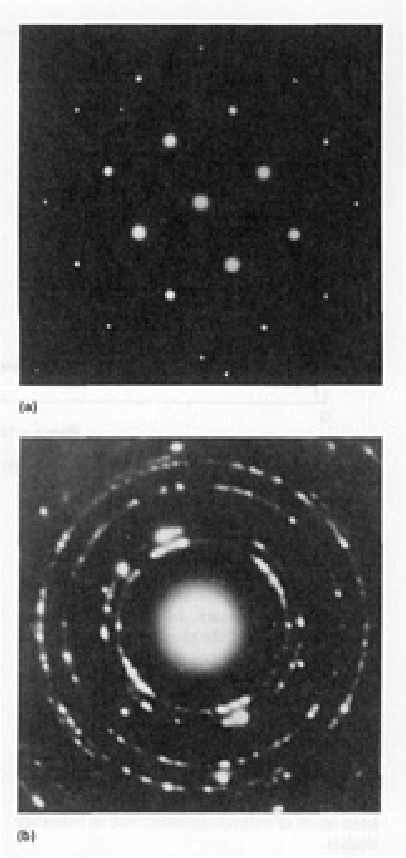


Fig. 15 Selected-area diffraction patterns. (a) Single-crystallite aluminum, (100) orientation. (b) Many crystallites near abraded copper surface. Source: Ref 27

Due to spherical aberration of the objective lens, as well as any focusing error, material slightly outside the region defined by the area-selecting aperture also contributes to the diffraction pattern. In practice, this error is not a problem unless the diffracting region must be known to high precision or the region to be selected is very small. Then the error limits the minimum selectable area to a diameter of about 1 or 0.5 μm at 100 or 300 kV, respectively.

For smaller areas, one of the so-called microdiffraction modes must be used (Ref 19). One method employs very small condenser apertures to limit the size of the beam. A second method requires an electron lens system similar to that in instruments capable of STEM operation to produce a very narrow and intense but still parallel beam of electrons. In both cases, the illuminating beam itself defines the area from which the diffraction pattern is obtained. This makes it possible to obtain spot patterns similar to those in the SAD mode from a specimen area less than one tenth of that possible in SAD.

In addition to spot or Laue-type patterns obtained by diffraction of the incident electron beam, a pattern of lines or bands referred to as a Kikuchi pattern may be observed (Fig. 16). To obtain a well-defined Kikuchi pattern, the illuminated specimen region must be relatively free of strain and not too thin. The lines occur because of the diffraction of electrons that have been subject to previous inelastic scattering within the specimen. As pointed out earlier, only inelastically scattered electrons that have lost relatively little energy can be brought to focus and are thus able to contribute to the pattern. In the absence of diffraction, these inelastically scattered electrons would form a uniform background with diminishing intensity at increasing distance from the central beam. Diffraction of these electrons by a given set of planes produces bright lines. Because these diffracted electrons are now lost from the background, there is an accompanying dark or deficiency line at 2θ distance from the bright line according to Bragg's law. Because the electrons that give rise to the Kikuchi pattern originate within the specimen, when the specimen is tilted the lines move or sweep across the field of view as though they were attached to the specimen. From a practical point of view, the Kikuchi pattern gives a more precise indication of the specimen orientation than the spot pattern. Most importantly, it allows the specimen to be oriented exactly as required for imaging and for the analysis of dislocations and other lattice defects.

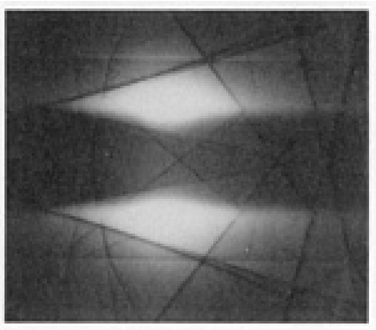


Fig. 16 Kikuchi diffraction pattern

If the specimen is illuminated with a focused converging beam instead of a parallel beam of electrons, a convergent-beam electron diffraction (CBED) pattern is formed. The size of the diffraction spots is increased substantially and internal contrast is visible (Fig. 17). If the illuminated specimen region is relatively free of defects and strain, "pendellosung" fringes may be observed within the spots. The spacing of these fringes can be used to determine the specimen thickness at the beam location (Ref 50).

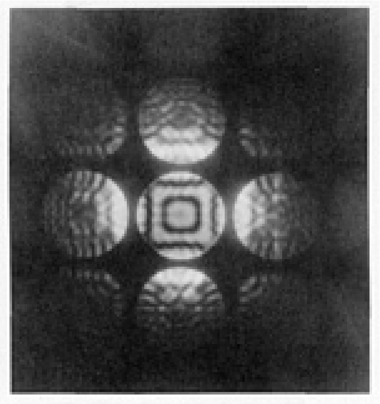


Fig. 17 CBED pattern for aluminum with (100) orientation

The CBED technique is extremely powerful. Because a focused probe is used, only the illuminated region contributes to the diffraction pattern; thus, this also qualifies as a microdiffraction mode. Also of great significance is the fact that the CBED technique allows the determination of crystal symmetry and crystal structure (Ref 19). By comparison, only limited symmetry information can be determined from SAD patterns. The procedure involves tilting the specimen so that the beam is along a low-index zone axis. From the detailed symmetry of such zone axis patterns (ZAPs), crystal structure information is obtained. The CBED method, coupled with compositional analysis by means of EDS and EELS, provides an extremely powerful method for the complete identification of small particles, precipitates, and microconstituents in a variety of materials.

To the three modes of diffraction (SAD, microdiffraction, and CBED) described above, reflection electron diffraction and several STEM modes of diffraction may be added. Reflection diffraction requires a smooth, nearly flat surface and a special holder or mount that allow the specimen to be oriented with its surface nearly parallel to the beam. In this case, the specimen need not be thin, because only the surface is involved. Reflection diffraction can be used to analyze thin films or to determine the orientation of a near-surface layer. Several different diffraction modes are available in STEM and will be discussed in a later section.

Image Contrast

Contrast in a TEM image can arise from almost any variation in the specimen material being examined. In general, these can be categorized as differences or variations in:

- Thickness
- Crystal structure
- Defects in the crystalline structure
- Crystal orientation
- Composition
- Strain
- Magnetic properties

More specifically, the observed contrast may be associated with different phases, which may be present in the form of precipitates, inclusions, atomic clusters, or embedded particles; lattice defects, such as dislocations, stacking faults, domain boundaries, and grain boundaries; strain associated with bending; crystal lattice planes and crystal structure; individual atoms under certain conditions; overlapping crystal that produce moiré patterns; and magnetic domains.

The size of the observable features can range from millimeters (that is, the dimensions of the viewable specimen) when the microscope is operated at its lowest magnification to the atomic scale at the highest magnifications. The origin of the contrast is often quite complex. A detailed quantitative explanation of the contrast lies very much in the realm of solid-state physics and quantum mechanics. Fortunately, most of the analyses have been worked out, and once the principles are understood, much of the interpretation can be carried out without recourse to theoretical details and computations. Examples of such analyses include the determination of dislocation Burgers vectors and the characterization of stacking faults. Exceptions where detailed compilations are required to exist, however. In particular, the analysis of high-resolution images to determine the details of atomic structure usually must be carried out in conjunction with computer simulation to verify any hypothesized interpretation (Ref 51).

Type of Contrast

In general, there are two ways in which contrast can be formed in the image: amplitude contrast and phase contrast.

Amplitude contrast arises from the variation in the number of electrons that leave the back surface of the specimen and reach the image. This is determined not only by scattering processes in the specimen but also by the fact that only electrons within a small angular range of the electron-optic axis can be brought to focus. In fact, the objective aperture is used to determine this angle and therefore to control contrast. If only the primary beam is allowed to pass through the aperture, a bright-field image results. Diffracted electrons are stopped by the objective aperture and do not contribute directly to the image. If more electrons are diffracted from one region of the specimen than from another, the former region will appear dark in the bright-field image. Similarly, inelastically scattered electrons that fall outside the objective aperture do not contribute, and a region where inelastic scattering is strong will appear dark. Contrast that is produced by inelastic scattering of this type is referred to as absorption contrast. Those inelastically scattered electrons that do enter the aperture produce a diffuse background.

If only diffracted electrons are allowed to pass through the objective aperture, a dark-field image results. Specimen regions contributing most to the diffracted beam will appear bright in the image. Small precipitates, for example, can be identified in this way. Ray diagrams illustrating the bright- and dark-field modes are shown in Fig. 14.

Phase Contrast. A phase contrast image results from the interference between two or more transmitted beams that have left the specimen. These interfering beams must of course pass through the objective aperture. The most notable application of this mode of imaging is in the formation of lattice images and crystal structure images (Ref 51).

Diffraction Contrast

Crystalline materials probably constitute the majority of the specimens of interest to tribologists. Diffraction contrast is the primary means by which the microstructures of crystalline materials are revealed. An example of diffraction contrast is shown in Fig. 18. The specimen is a polycrystalline aluminum alloy thinned by electropolishing. A bright-field image is shown. Because of differences in crystallographic orientation, there are in some cases marked differences in contrast

among the grains. By tilting the specimen--that is, by changing the diffracting conditions--grain that are dark could be made to appear bright and vice versa.

Fig. 18 Bright-field image of a polycrystalline aluminum alloy

Alternatively, one of the diffracted beams could have been selected for imaging to obtain a dark-field image. This is best accomplished by tilting the illuminating beam (Fig. 14) while observing the diffraction pattern. The diffraction spot of interest is brought to the center of the screen. An objective aperture is then introduced, excluding other spots, and the instrument is switched to the imaging mode. Another alternative would be simply to move the objective aperture to the diffraction spot of interest, but this results in a blurred image because of the spherical aberration of the objective lens. In any case, grains that do not contribute to the selected diffraction spot will appear dark.

Diffraction contrast provides the means of revealing and characterizing lattice defects, such as dislocations and stacking faults, and for observing precipitates and other second-phase particles. To explain the origin of the observed contrast, consideration must be given to the manner in which electrons propagate through the specimen material. Fundamentally, this involves the determination of the wave function of the electron as it encounters the potential field associated with the assembly of atoms constituting the specimen. The wave function is obtained by solving Schrödinger's equation. This approach is referred to as the dynamic theory. The theory is considerably simplified if only two beams are considered--specifically, the primary beam and one diffracted beam. Realization of this condition is accomplished by tilting the specimen so that only the primary beam and one diffracted beam are strongly excited in the diffraction pattern.

At high accelerating voltages (>300 kV) and for materials with large lattice spacings, it may not be possible to obtain a single strongly excited diffracted beam without strongly exciting other reflections in a systematic row (for example, exciting the 111 reflection without exciting the 222 and 333 reflections). Accurate computation of image contrast under these circumstances requires the so-called many-beam theory.

The kinematic theory, which is much simpler than the dynamic theory, can sometimes be employed to determine contrast. Strictly speaking, the kinematic theory is valid only for very thin specimens and when the diffracted beams is weak compared with the primary beam. The theory is essentially a geometric optics approach. The scattering of electrons by an array of atoms is treated like the scattering of light by a grating. A number of interference and diffraction effects can be demonstrated by this approach.

An important result of the two-beam dynamical theory is the relationship describing the intensity of the diffracted beam, *I*(D), and of the transmitted primary beam, *I*(T):

$$\begin{aligned} I(\mathrm{D}) &= 1 - I(\mathrm{T}) \\ &= \frac{1}{1+w^2} \sin^2 \left(\pi \sqrt{1+w^2}\frac{t}{\xi_g} \right) \end{aligned} \quad \textbf{(Eq 3)}$$

where t is the thickness of the specimen normal to the beam ξ_g is the extinction distance, and $w = s\xi_g$. The parameter s is a measure of the deviation from the exact orientation for Bragg diffraction: $s = 0$ indicates that the specimen is oriented so that the Bragg condition (Eq 2) is exactly satisfied. Also, the total transmitted intensity, $I(T) + I(D)$, is normalized to 1.

The extinction distance, ξ, is an extremely important parameter in this relationship. It accounts for the periodic variation in intensity that occurs as a function of specimen thickness, an effect that is most obviously demonstrated by the bright and dark fringes seen at the edge of thin, wedge-shaped specimens. Physically, this effect is explained by the fact that rediffraction occurs between the transmitted and diffracted beams. Thus, intensity is built up in the diffracted beam, reaches a peak, and is then returned to the transmitted beam, and so forth. The distance between peaks is the extinction distance. The extinction distance also affects the contrast seen at dislocations, stacking faults, bend contours, grain boundaries, and second-phase particles.

To describe the contrast that occurs at various features (dislocations, stacking faults, precipitates, and so on), the column approximation is usually employed (Ref 13). As illustrated in Fig. 19, a slab of material containing the feature of interest is divided into equal-size columns parallel to the beam. Each column is assumed to be large enough in cross section to contain both the transmitted and diffracted beams, but to be sufficiently small that variations caused by the feature across the column can be ignored. The amplitudes of the coupled transmitted and diffracted waves are integrated down the column. The contributions of all columns then gives the image.

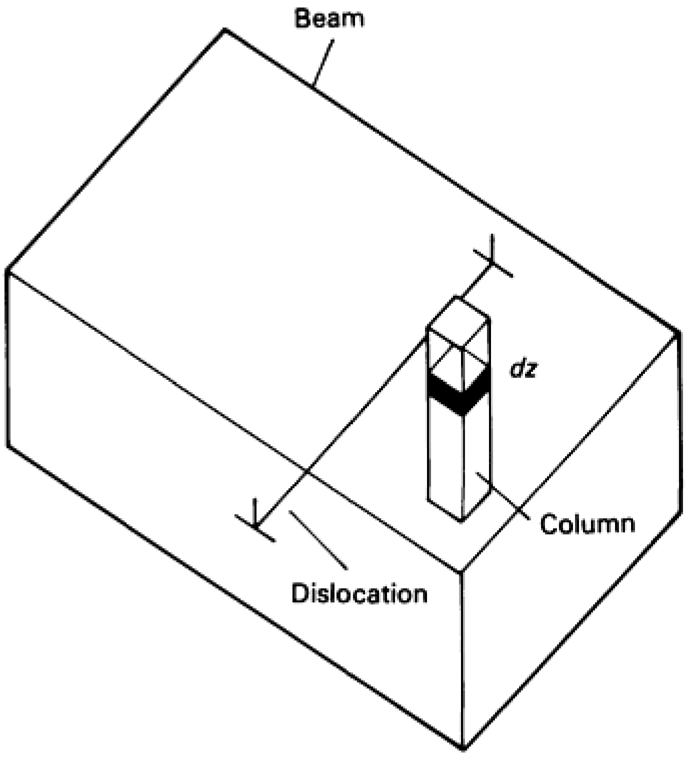


Fig. 19 Illustration of column approximation used to calculate contrast near lattice defects

Maps of intensity associated with various defects have been determined by computer methods (Ref 19), although in many cases one or a few computed profiles of intensity across the feature are suitable to establish the required identification.

For many metals and alloys where the nature of the possible lattice defects is already known, and even in some cases where it may not be, such detailed computations are not required. Rather, all that are needed for identification are the relatively simple relationships between the diffracting condition, which is determined by the orientation of the specimen, and parameters characterizing the feature, such as the possible Burgers vector of a dislocation or the displacement vector of a stacking fault.

Figure 20 illustrates schematically the means by which diffraction contrast can give rise to the image of a distortion, in this case an edge dislocation. Strain caused by the presence of the dislocation is accommodated by the local tilting of the lattice planes. Planes perpendicular to the Burgers vector are most affected. It is assumed that the specimen has already been tilted to a two-beam condition and that these planes are the diffracting planes. Furthermore, it is assumed that the Bragg condition is not quite satisfied for planes distant from and therefore little affected by, the presence of the dislocation. This is tantamount to the parameter *s* in Eq 3 having a value slightly different from zero. As a consequence, less than the maximum intensity will be carried by the diffracted beam, except in the vicinity of the dislocation, where planes, at least on one side of the dislocation, may be oriented for exact Bragg diffraction . For these planes, $s = 0$ and the maximum intensity can appear in the diffracted beam. As a result, the dislocation will be displayed, at least nominally, as a dark line of contrast under bright-field conditions and as a bright line under dark-field conditions.

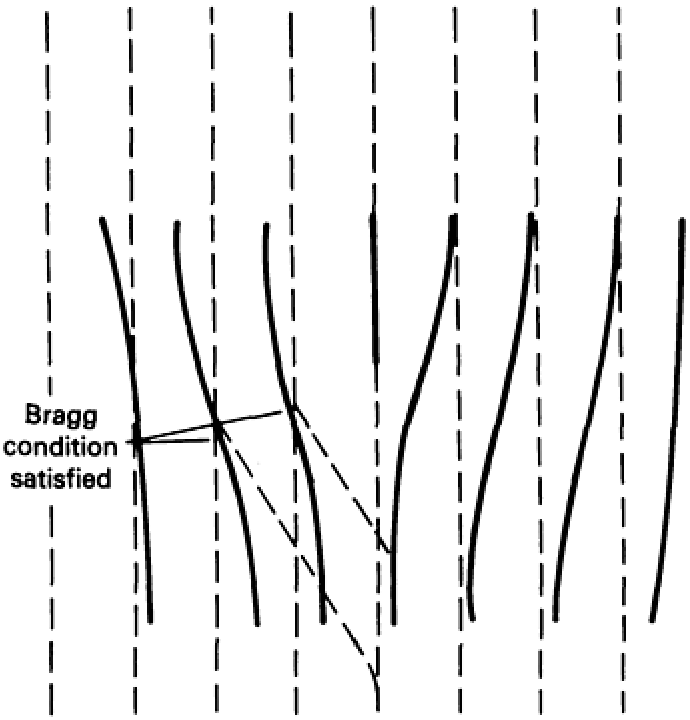


Fig. 20 Schematic showing origin of diffraction contrast from edge dislocation

If the specimen is oriented so that only planes parallel to the Burgers vector are diffracting, almost no contrast will be seen, because the diffracting planes are now only slightly affected by the presence of the dislocation. This result is expressed by the relation:

$$\mathbf{g} \cdot \mathbf{b} = 0 \qquad \textbf{(Eq 4)}$$

where **g** is a vector perpendicular to the diffracting lattice planes and **b** is the Burgers vector of the dislocation. Thus, by experimenting with different diffracting conditions (that is, by selecting various **g** vectors), the direction of the Burgers vector of the dislocation can be determined.

Under conventional two-beam diffraction contrast conditions, where **g** refers to the lowest index set of planes in the series (the diffraction spot in the systematic row that is closest to the center of the pattern), the images of dislocations are 10 to 20 nm in width. Much narrower images and therefore greatly improved resolution can be achieved by utilizing a higher-order **g** vector. Thus, for example, it is possible to resolve individual partial dislocations in copper that are separated by approximately 2.3 nm. This method is referred to as weak beam imaging (Ref 14, 19).

An example of dislocations observed in a worn specimen is shown in Fig. 21. The specimen is relatively pure copper that was subjected to solid particle erosion. The dislocations are arranged in a cell structure, and, in fact, the size of the cells is

an indication of the amount of strain in the material. In the study from which this example was taken, cross sections were prepared perpendicular to the worn surface and the cell size measured as a function of distance below the surface. From this, the relationship between the amount of strain and depth below the worn surface was determined. Additional discussion of the relationship between cell size and strain at worn surfaces, together with illustrative TEM images, can be found in Ref 35.

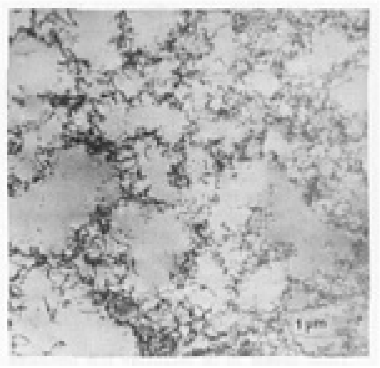


Fig. 21 TEM micrograph of cross section through an eroded copper surface showing dislocation cell structure. Source: Ref 52

Deformation and the associated presence of dislocations also play an important role in the response of nonmetallic materials to tribological contacts. In sliding wear experiments, the dislocation structure near the worn surface of an alumina (Al_2O_3) doped with MgO) specimen exhibited a transition from low wear to high, severe wear (Ref 53). The transition was attributed to strain-induced cracking at grain boundaries. The strain arose from the accumulation of dislocations during the low wear regime. The specimen from which Fig. 22 was obtained has been exposed to prolonged wear in the low wear regime. Prior to exposure to wear, the grains were essentially free of dislocations.

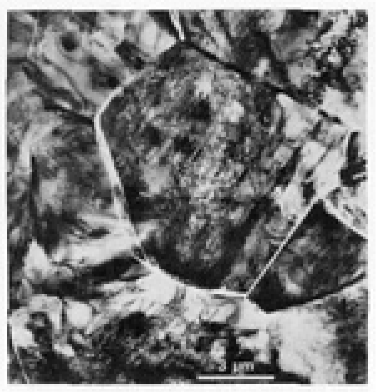


Fig. 22 Accumulation of dislocations near grain boundaries caused by sliding contact. The material is Al_2O_3 doped with MgO. Source: Ref 53

In addition to imaging dislocations, other lattice defects, such as stacking faults, grain boundaries, and boundaries between ordered and disordered regions, can be revealed in the TEM. Many examples, together with detailed discussions of the analysis schemes for characterizing these defects, can be found in Ref 13, 14, 15, 16.

An investigation of wear will usually involve the study of one or more materials, and each material may consist of several phases. Identification and characterization of these materials and phases can play a critical role in understanding the wear process and perhaps in selecting or developing improved materials. Commercial alloys typically consist of more than one phase. Depending on composition and processing method, the phase morphology can range from relatively large grains to extremely small, coherent precipitates. All are subject to change under the influence of a tribological contact. The

thermal, mechanical, and chemical effects associated with the tribological contact may modify an original structure or create an entirely new structure. The conditions in the contact can be assessed on the basis of the presence of a particular phase, which might occur only if, for example, a specific temperature had or had not been exceeded. Oxide and other films formed by chemical reaction of the surface with the surrounding environment and perhaps a lubricant require characterization. Wear debris may form a mechanically alloyed layer on the surface (Ref 35, 54) and should be analyzed.

An interesting example where TEM was used together with SEM and light optical microscopy to reveal changes in the microstructure of AISI 52100 bearing steel as a result of rolling contact fatigue can be found in an investigation by Osterlund and Vingsbo (Ref 55). The microstructure of AISI 52100 when heat treated for bearing applications consists primarily of tempered lath martensite, with a small amount of retained austenite and a dispersion of small (~1 μm diam) carbides throughout. Depending on load, after about 10^7 cycles, the martensite begins to decay in the region below the contact surface that experiences the maximum Hertzian stress. Evidence of the change can be observed on carefully etched sections by light optical and SEM examination. Detailed characterization of the changes, however, requires TEM analysis. The change essentially involves the transformation of martensite to ferrite and the formation of carbide precipitates (Ref 55). Excellent TEM micrographs illustrating the decay process can also be found in an earlier paper by Swahn *et al.* (Ref 56).

Phase Contrast Imaging

Phase contrast imaging has become an increasingly important imaging mode. The primary application has probably been the study of the atomic-scale crystal structure of electronic materials and devices, but the technique is also being used extensively to study structural ceramics, composites, cermets, intermetallics, and metal alloys. The field of electron microscopy devoted to the study of structure on the atomic scale is generally referred to as high-resolution electron microscopy (HREM). For a full discussion of HREM, see Ref 51.

High-resolution studies require a suitable TEM, such as intermediate-voltage instruments capable of 0.17 to 0.2 nm resolution or high-voltage instruments that provide even better resolution. Considerable operator skill as well as a sound knowledge of the basic principles involved are prerequisites for most high-resolution studies. Moreover, specimen requirements are quite stringent. For crystal structure images, the thickness should generally be no greater than about 40 nm.

Obtaining a lattice image requires that at least two beams (the primary and one diffracted beam, for example) enter the objective aperture. Interference due to the phase difference between the two beams results in periodic fringes corresponding to the lattice planes. Phase changes also are introduced by the spherical aberration of the objective lens and objective lens defocus. Adjusting the focus of the objective lens allows an optimum image to be obtained; this is referred to as the Scherzer focus. The capability of the objective lens to produce this type of contrast is expressed in terms of the so-called contrast transfer function of the lens.

When reflections from different, noncoplanar sets of planes enter the objective aperture and interfere, the positions of atom rows can be imaged. Interpretation of the actual crystal structure, that is, assigning specific atoms to the rows, is complicated by the fact that the contrast is sensitive to a number of different factors, including the specimen thickness and the contrast transfer function of the objective lens. In practice, a structure is hypothesized and a simulated image is computed by employing the appropriate variables. When the computer-generated and observed images are the same, it can be assumed that the hypothesized structure is valid.

HREM can be used to study the structure of triboreaction films and solid lubricants and, of course, the detailed changes in tribocomponent materials. Martin *et al.* (Ref 57) used HREM to observe lattice fringes of iron-rich crystallites in triboreaction film debris fragments generated by sliding AISI 52100 steel against cast iron lubricated with oil containing the additive zinc diisopropyldithiophosphate. In another investigation, Ganapathi and Rigney (Ref 58) studied the mechanically mixed layer that formed on a copper block slid on type 440 stainless steel in argon. TEM showed that the structure at the surface consisted of extremely small "nanosize" grains. High-resolution lattice images were used to study the detailed structure of these nanocrystalline grains and the associated grain boundaries.

EDS Analysis in the TEM

Energy-dispersive x-ray analysis is conjunction with TEM is conducted in essentially the same way as with SEM. The same types of detector and analysis systems are employed with both microscopes. The main differences lie in the much higher accelerating voltages and thin, electron-transparent specimens that are typical of TEM applications. This

combination can have significant advantages. For sufficiently thin specimens, absorption and fluorescence corrections are not necessary, resulting in a considerable simplification in the quantitative determination of composition. The maximum allowable specimen thickness for this condition to hold depends on the material and the accelerating voltage (Ref 51). The maximum thickness is increased for low-atomic-number materials and higher accelerating voltages.

Thin specimens also result in a significant improvement in spatial resolution. In analyzing a bulk sample in the SEM, the volume from which x-rays are collected is on the order of 1 μm in diameter. In the TEM, with a sufficiently thin specimen, the volume may be little greater than that established by the beam diameter and the specimen thickness. With a beam diameter that may be as small as 1 nm and a specimen thickness as small as a few nm, this represents an extremely small volume indeed. As a result of this high spatial resolution, accurate compositional analyses can be obtained for small precipitates and other second-phase particles. Also, composition profiles in the vicinity of interfaces and grain boundaries can be determined precisely.

Electron Energy Loss Spectrometry

Electron energy loss spectrometry in the TEM is most important in connection with the analysis of light elements, where the application of EDS is limited, and in the determination of atomic bonding information, which is not available from x-ray spectra. Thus, EELS is capable not only of determining the elemental composition of a specimen but also of assessing its full chemical identity. A thorough discussion of EELS can be found in Ref 18, 19, and 59.

An energy loss spectrum is schematically depicted in Fig. 23, where the difference between the energy of the incident and scattered electrons--that is, the amount of energy lost--is plotted as a function of intensity. The spectrum, as illustrated in Fig. 23, is usually separated into two parts: the low-loss region and the high-loss region. The low-loss region ranges from approximately 0 eV (no energy loss) to about 50 eV, and the high-loss region extends beyond 50 eV. On average, the intensity falls off rapidly with increasing energy loss and becomes very small in the high-loss region. In order to show the details of both the low- and high-loss regions in the same graph, as is customary, the high-loss region is displayed at an increased gain.

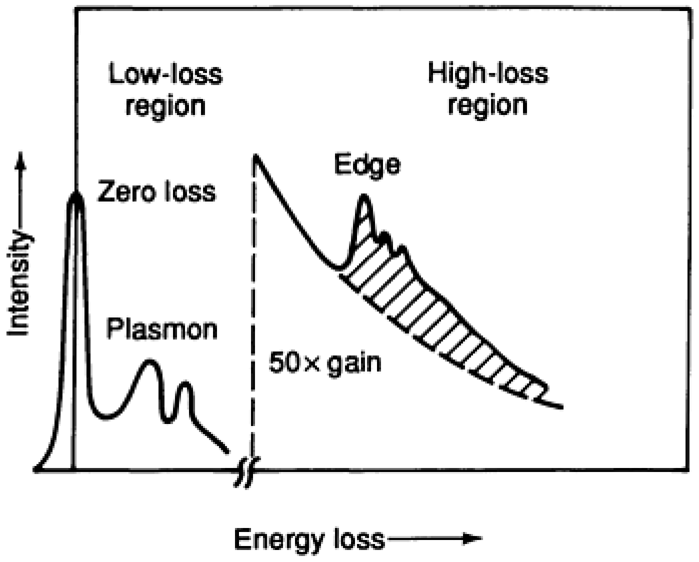


Fig. 23 Schematic illustration of EELS spectrum

In Fig. 23, the first and largest peak in the low-loss region is the zero-loss peak. This peak consists primarily of incident beam electrons that have passed through the specimen without being scattered and of electrons that have been scattered without losing energy (elastically scattered electrons). The incident electrons beam is not monochromatic, but has a small energy spread. This, together with the finite resolution of the spectrometer, results in the measurable width of the zero-loss peak.

The relatively broad peaks following the zero-loss peak result from collective, or plasmon, interactions with conduction and/or valence band electrons. By comparing the integrated intensity of these plasmon peaks with that of the zero-loss peak, it is possible to obtain information on specimen thickness.

The high-loss region of the spectrum is by far the most important. It is from this region that most of the information about specimen chemistry is obtained. The ejection of an inner-shell atomic electron (that is, ionization of the atom) results in a characteristic absorption edge that can be used to determine the elemental composition of the specimen. (It may be recalled that the emission of x-rays during decay of the ionized atom is the basis of EDS analysis.) In addition, the shape immediately following the onset of the edge, the so-called energy-loss near-edge structure (ELNES), provides information about the atomic environment around the atomic species responsible for the edge.

The most commonly used EELS spectrometer is positioned on the optical axis at the base of the microscope, below the viewing screen. Electrons passing through the entrance slit of the spectrometer are energy dispersed by a magnetic field. The dispersed electrons can be detected either serially by being scanned across a single detector or in parallel by an array of detectors. The latter method takes much less time to acquire a spectrum. The signal associated with the detected electrons is amplified and sent to an analysis system for processing and display. The same analysis system used for EDS can be used for EELS.

By positioning a particular area in the specimen image over the entrance aperture of the spectrometer, the energy loss spectrum from that area and the features within can be collected and analyzed. The size of the analyzed area is defined by the entrance aperture and can be controlled by adjusting the image magnification. Alternatively, a diffraction spot can be placed over the entrance aperture and the entire specimen region contribution to that spot subjected to analysis. In the STEM mode of operation, it is possible to map the composition of the specimen or to obtain an energy-filtered image.

There are several limitations to EELS analysis. The specimen must be quite thin, generally less than 50 nm, to avoid multiple scattering events as the electron passes through the material. Multiple scattering results in a rapid increase in background level, which eventually overcomes the spectral features of interest. In addition, the application of EELS to specimens containing more than a few elements may be severely limited because of extensive peak overlap. Also, the quantitative determination of concentrations is less well developed in EELS than in EDS.

There appear to be relatively few instances where EELS has been applied directly to the study of wear. One example concerning the analysis of boundary lubrication films can be found in Ref 57.

Scanning Transmission Electron Microscopy

Many TEMs are designed to operate in the STEM mode as well as in the stationary-beam CTEM mode. The desired mode of operation is selected according to the requirements of the particular investigation and to some extent by the preferences of the operator. The conventional operating mode is probably most convenient for routine examination of microstructure, diffraction contrast studies of defect structures, crystal structure studies using electron diffraction, and for EDS and EELS analyses of specific features. Also, only when the microscope is equipped with a field emission source does resolution in the STEM mode approach that in the CTEM mode. Thus, the CTEM mode is generally employed for high-resolution studies.

The STEM mode has three extremely important and the useful capabilities not available in the CTEM mode, or available only to a limited extent. First, virtually any emitted signal for which a detector is available can be used to produce an image in the STEM mode. Second, the electronic signal is easily processed and enhanced to control contrast and brightness for suitable viewing and photography. Third, digital beam control and image digitization are relatively easy steps, allowing the direct application of computer processing and analysis to the image. For example, with digital beam control, all particles meeting a certain size or shape criterion might be selected for further compositional analysis, which would then proceed automatically.

Once in the STEM mode, the microscope functions essentially as an SEM. CRTs are available for visual observation and photography. Alternatively, with a modern digital instrument, a computer screen replaces the visual CRT. Bright- and dark-field detectors are basic equipment in the STEM system. According to the principle of reciprocity (Ref 19). STEM bright-field and dark-field transmission images are similar to CTEM bright- and dark-field images, or at least can be made similar by the selection of appropriate aperture angles. Thus, the rules regarding image contrast interpretation in CTEM carry over to STEM images.

STEM offers several different methods for obtaining a diffraction pattern. The incident beam can be rocked about a point on the specimen surface, whereby the diffraction pattern is scanned across the detector and displayed on the CRT monitor. Alternatively, the incident beam can be kept stationary and the transmitted beam scanned over the detector, or both the incident and transmitted beams can be deflected in synchronism to generate the diffraction pattern. By using the image analysis capabilities inherent in the STEM mode, analysis of the diffraction pattern can be carried out directly in conjunction with appropriate computer software.

In addition to bright- and dark-field detectors, an array of other detectors is also available, including secondary electron, backscattered electron, and emitted light (cathodoluminescence) detectors. These detectors are used to examine the specimen surface and allow direct comparison of surface topography and internal structure revealed by transmission detectors. When equipped with EDS and EELS systems, signals detected by these systems may also be used for image generation to produce composition maps. As already discussed in the sections on EDS and EELS in the TEM, the thin, electron-transparent specimen results in a significant improvement in spatial resolution compared with bulk analysis in the SEM by EDS.

Summary and Future Outlook

A detailed understanding of wear processes and mechanisms will, in general, require comprehensive information on the response of the materials involved. Surface topography, reaction films, subsurface microstructure, and wear debris all require characterization. The SEM has become the primary tool in such investigations. Ease of operation, an extremely large magnification range, high resolution, great depth of field, and lack of restrictive specimen preparation requirements, together with the ability to determine elemental composition with associated x-ray analysis facilities, are responsible for this status. With the continued improvement in capabilities, particularly in connection with the development of computer-based instruments with more extensive compositional mapping and quantitative image analysis capabilities, it seems likely that the SEM will be of even greater value in tribological studies in the future.

The TEM has not been as extensively used in the study of wear as the SEM, but nevertheless has made significant contributions to the understanding of wear processes. Specimen preparation for TEM analyses, in general, requires considerable effort. Image interpretation is more demanding and may require a substantial understanding of the physics of electron-specimen interactions. The TEM is capable of providing detailed information on the microstructure and composition of materials, even at the atomic scale of resolution in some cases. The nature and arrangement of dislocations resulting from tribological contact, thermal and stress-induced phase changes, the composition and microstructure of triboreaction films, and the properties of wear debris can all be characterized by means of TEM.

References

1. K.C. Ludema and R.G. Bayer, Ed., *Wear of Materials--1991,* American Society of Mechanical Engineers, 1991
2. R. Devanathan and P. Clayton, Rolling/Sliding Wear Behavior of Three Bainitic Steels, *Wear of Materials--1991,* K.C. Ludema and R.G. Bayer, Ed., American Society of Mechanical Engineers, 1991, p 91
3. I.L. Singer, S. Fayeulle, and P.D. Ehni, Friction and Wear Behavior of TiN in Air: The Chemistry of Transfer Films and Debris Formation, *Wear of Materials--1991,* K.C. Ludema and R.G. Bayer, Ed., American Society of Mechanical Engineers, 1991, p 229
4. W. Bundschuh and K.-H. Zum Gahr, Influence of Porosity on Friction and Sliding Wear of TZP-Zirconia, *Wear of Materials--1991,* K.C. Ludema and R.G. Bayer, Ed., American Society of Mechanical Engineers, 1991, p 319
5. L.K. Ives, J.S. Harris, and M.B. Peterson, Evaluation of a New Wear Resistant Additive--$SbSbS_4$, *Wear of Materials--1983,* American Society of Mechanical Engineers, 1983, p 507
6. P.J. Shuff and L.J. Clarke, Imaging of Lubricating Oil Insolubles, *Tribology Int.,* Vol 24, 1991, p 381-387
7. A.A. Milne and W.L. Cooke, On the Influence of Grease Structure on Boundary Lubrication, *Proceedings of the Institute of Mechanical Engineers Conference on Lubrication and Wear,* Institute of Mechanical Engineers, London, 1957, p 350
8. J.I. Goldstein, D.E. Newbury, P. Echlin, D.C. Joy, C.E. Fiori, and E. Lifshin, *Scanning Electron Microscopy and X-Ray Microanalysis,* Plenum Press, 1981

9. D.E. Newbury, D.C. Joy, P. Echlin, C.E. Fiori, and J.I. Goldstein, *Advanced Scanning Electron Microscopy and Analysis,* Plenum Press, 1986
10. C.E. Lyman, D.E. Newbury, J.I. Goldstein, D.B. Williams, A.D. Romig, Jr., J.T. Armstrong, P. Echlin, C.E. Fiori, D.C. Joy, E. Lifshin, and K.-R. Peters, *Scanning Electron Microscopy, X-Ray Microanalysis, and Analytical Electron Microscopy: A Laboratory Workbook,* Plenum Press, 1990
11. L. Reimer, *Scanning Electron Microscopy,* Springer-Verlag, Berlin, 1985
12. B.L. Gabriel, *SEM: A User's Manual for Materials Science,* American Society for Metals, 1985
13. P.B. Hirsch, A. Howie, R.B. Nicholson, D.W. Pashley, and M.J. Whelan, *Electron Microscopy of Thin Crystals,* 1977
14. G. Thomas and M.J. Goringe, *Transmission Electron Microscopy of Materials,* John Wiley & Sons, 1979
15. L. Reimer, *Transmission Electron Microscopy,* 2nd ed., Springer-Verlag, Berlin, 1989
16. J.W. Edington, *Practical Electron Microscopy in Materials Science,* Van Nostrand Reinhold, 1976
17. D.C. Joy, A.D. Romig, Jr., and J.I. Goldstein, Ed., *Principles of Analytical Electron Microscopy,* Plenum Press, 1986
18. D.B. Williams, *Practical Analytical Electron Microscopy in Materials Science,* Electron Optics Publishing Group, 1984
19. J.J. Hren, J.I. Goldstein, and D.C. Joy, *Introduction to Analytical Electron Microscopy,* Plenum Press, 1979
20. M. von Heimendahl, *Electron Microscopy of Materials,* Academic Press, 1980
21. J.D. Verhoeven, Scanning Electron Microscopy, *Metals Handbook,* 9th ed., Vol 10, *Materials Characterization,* American Society for Metals, 1986, p 490
22. A.D. Romig, Jr., Analytic Transmission Electron Microscopy, *Metals Characterization,* American Society for Metals, 1986, p 429
23. P.J. Goodhew and F.J. Humphreys, *Electron Microscopy and Analysis,* Taylor and Francis, London, 1988
24. D.B. Williams and D.E. Newbury, Recent Advances in the Electron Microscopy of Materials, *Advances in Electronics and Electron Physics,* P.W. Hawkes, Ed., Academic Press, London, 1984, p 161-288
25. T.F.J. Quinn, *The Application of Modern Physical Techniques to Tribology,* Van Nostrand & Reinhold, 1971
26. T.F.J. Quinn, *Physical Analysis for Tribology,* Cambridge University Press, 1991
27. A.W. Ruff, L.K. Ives, and W.A. Glaeser, Characterization of Wear Surfaces and Wear Debris, *Fundamentals of Friction and Wear,* D.A. Rigney, Ed., American Society for Metals, 1980, p 235-289
28. W.A. Glaeser, Wear Experiments in the Scanning Electron Microscope, *Wear,* Vol 73, 1981, p 371
29. S.J. Calabrese, F.F. Ling, and S.F. Murray, Dynamic Wear Tests in the SEM, *ASLE Trans.,* Vol 26, 1983, p 455
30. W. Holzhauer and F.F. Ling, In-Situ SEM Study of Boundary Lubricated Contacts, *ASME Trans.,* Vol 31, 1987, p 359
31. T. Kayaba, K. Hokkirigawa, and K. Kato, Analysis of the Abrasive Wear Mechanism by Successive Observations of Wear Processes in a Scanning Electron Microscope, *Wear,* Vol 110, 1986, p 419
32. K Hokkirigawa and K. Kato, The Effect of Hardness on the Transition of the Abrasive Wear Mechanism of Steels, *Wear,* Vol 123, 1988, p 241
33. H. Kitsunai, N. Tsumaki, and K. Kato, Transitions of Microscopic Wear Mechanism for Cr_2O_3 Ceramic Coatings During Repeated Sliding Observed in an SEM-Tribosystem, *Wear of Materials--1991,* K.C. Ludema and R.G. Bayer, Ed., American Society of Mechanical Engineers, 1991, p 249
34. J.H. Dautzenberg and J.H. Zaat, Quantitative Determination of Deformation by Sliding Wear, *Wear,* Vol 23, 1973, p 9
35. D.A. Rigney, M.G.S. Naylor, R. Divakar, and L.K. Ives, Low Energy Dislocation Structures Caused by Sliding and Particle Impact, *Mater. Sci. Eng.,* Vol 81, 1986, p 409
36. D.A. Rigney and W.A. Glaeser, The Significance of Near Surface Microstructure in the Wear Process, *Wear,* Vol 48, 1978, p 241

37. A.W. Ruff, Characterization of Debris Particles Recovered From Wearing Systems, *Wear,* Vol 42, 1977, p 49
38. P.J. Goodhew, *Specimen Preparation for Transmission Electron Microscopy,* Oxford University Press, 1984
39. J.C. Bravman, R.M. Anderson, and M.L. McDonald, Ed., *Specimen Preparation for Transmission Electron Microscopy of Materials,* MRS Symp. Proc. Vol 115, Materials Research Society, 1988
40. R.M. Anderson, Ed., *Specimen Preparation for Transmission Electron Microscopy of Materials--II,* MRS Symp. Proc. Vol 199, Materials Research Society, 1990
41. I.S. Brammar and M.A. Dewey, *Specimen Preparation for Electron Metallography,* Blackwell, Oxford, 1996
42. J.C. Bravman and R. Sinclair, The Preparation of Cross-Section Specimens for Transmission Electron Microscopy, *J. Electron Microsc. Tech.,* vol 1, 1984, p 53
43. L.E. Samuels, *Metallographic Polishing by Mechanical Methods,* 3rd ed., American Society for Metals, 1982
44. R. Alani, Recent Advances in Ion Milling Techniques for TEM Specimen Preparation of Materials, *Electron Microscopy,* World Publishing, London, 1991, p 461
45. T. Allen, *Particle Size Measurement,* 3rd ed., Chapman and Hall, 1981, p 196
46. A.W. Ruff, Deformation Studies at Sliding Wear Tracks in Iron, *Wear,* Vol 40, 1976, p 59
47. K.F.J. Heinrich, *Electron Beam X-Ray Microanalysis,* Van Nostrand Reinhold, 1981
48. L.K. Ives, Abrasive Wear by Coal-Fueled Engine Particles, *Proceedings Corrosion Erosion-Wear of Materials at Elevated Temperatures,* A.V. Levy, Ed., National Association of Corrosion Engineers, 1991, p 29-1 to 29-20
49. P.J. Blau and C.D. Olson, An Application of Thermal Wave Microscopy to Research on the Sliding Wear Break-In Behavior of a Tarnished Cu-15 wt% Zn Alloy, *Wear of Materials--1985,* K.C. Ludema, Ed., American Society of Mechanical Engineers, 1985, p 424
50. P.M. Kelly, A. Jostsons, R.G. Blake, and J.G. Napier, The Determination of Foil Thickness by STEM, *Phys. Stat. Sol. A,* Vol 31, 1975, p 771
51. J.C.H. Spence, *Experimental High-Resolution Electron Microscopy,* 2nd ed., Oxford University Press, 1988
52. L.K. Ives and A.W. Ruff, Electron Microscopy Study of Erosion Damage in Copper, *Erosion: Prevention and Useful Application,* STP 664, ASTM, 1979, p 5-33
53. S.-J. Cho, H. Moon, B.J. Hockey, and S.M. Hsu, The Transition From Mild to Severe Wear in Alumina During Sliding, *Act. Metall. Mater.,* Vol 40, 1992, p 185
54. P. Heilmann, J. Don, T.C. Sun, D.A. Rigney, and W.A. Glaeser, Sliding Wear and Transfer, *Wear,* Vol 91, 1983, p 171
55. R. Osterlund and O. Vingsbo, Phase Changes in Fatigued Ball Bearings, *Metall. Trans. A,* Vol 11A, 1980, p 701
56. H. Swahn, P.C. Becker, and O. Vingsbo, Martensite Decay During Rolling Contact Fatigue in Ball Bearings, *Metall. Trans. A,* Vol 7A, 1976, p 1099
57. J.M. Martin, J.L. Mansot, I. Berbezier, and M. Belin, Microstructural Aspects of Lubricated Mild Wear With Zinc Dialkyl-dithiophosphate, *Wear,* Vol 107, 1986, p 355
58. S.K. Ganapathi and D.A. Rigney, An HREM Study of the Nanocrystalline Material Produced by Sliding Wear Processes, *Scr. Metall. Mater.,* Vol 24, 1990, p 1675
59. R.F. Egerton, *Electron Energy-Loss Spectroscopy in the Electron Microscope,* Plenum Press, 1986

Scanning Tunneling Microscopy

Yip-Wah Chung and T.S. Sriram, Northwestern University

Introduction

THE SCANNING TUNNELING MICROSCOPE (STM) represents a microscopy technique that enables high-resolution imaging of surfaces over a relatively large dynamic range, both in the horizontal and vertical directions. Even more significantly, this high resolution is achieved in vacuum, air, and liquid environments, which makes this new technique convenient to use for practical specimens.

This article describes the STM technique in the context of surface topography measurements, although applications in other areas are described, as well. To bring the subject into perspective, its history is described first, followed by a brief discussion of its physical basis. The technical details of the design and operation of the microscope are then discussed, followed by several examples of its applications. Finally, limitations and variants of the basic STM are described.

Historical Perspective

The theory of quantum mechanical tunneling was first developed in the early 1920s. An early triumph of this theory was its correct explanation of the strong dependence of the half-lives of radioactive nuclides on kinetic energies of emitted α particles. Leo Esaki was the first to exploit tunneling in a real solid-state device while working at the IBM Yorktown Heights Thomas J. Watson Research Center. In 1972, he received the Nobel Prize in physics for this invention. At about that time, Russell Young and his coworkers at the National Bureau of Standards (now the National Institute of Standards and Technology) described an instrument called the topografiner (Ref 1, 2), which in many ways is the predecessor of the modern STM. The topografiner operates in the field emission mode, with a horizontal resolution of about 20 nm and a vertical resolution of about 0.3 nm. Vibration and tip instabilities are the probable reasons that resolution was somewhat limited at this stage.

The concept of the STM was first described in a patent disclosure in mid-1979 by Heinrich Rohrer and his research staff assistant, Gerd Binnig. The next few years were spent solving problems related to vibration isolation and both coarse and fine motion control in all three axes. Their first success occurred in 1981, when images resolving monatomic steps on the surface of a calcium-iridium-tin crystal were obtained. However, the corresponding paper was rejected, because one referee found it "not interesting enough." What caught the attention of the scientific community in 1982 was the paper written by these researchers that described their success in imaging the famous Si(111)-(7 × 7) surface (Ref 3). Although this surface structure was first observed in the 1960s, its complexity precluded a structural solution until this work in 1982. In 1986, Binnig and Rohrer were awarded the Nobel Prize in physics for their STM work.

Since then, many STMs have been built and developed around the world. Commercial versions that operate both in air and in ultrahigh vacuum are also becoming available. Variants, such as the atomic force microscope and the scanning capacitance microscope, the expand the functionalities of the basic STM have also been developed. Applications cover many disciplines in physics, chemistry, biology, materials science, and engineering. Reports of using the STM for additive and subtractive lithography in the nanometer scale are emerging (Ref 4, 5, 6).

Principle of STM Imaging

Consider a sharp conducting tip brought to within 1 nm (0.04 μin.) of a specimen surface (Fig. 1). Typically, a bias ranging from 0.01 to 1 V is applied between the tip and the specimen. Under these conditions, the tip-surface spacing, *s*, is sufficiently small that electrons can tunnel from the tip to the specimen or vice versa. As a result, a current, *I*, flows across this gap and can be shown to vary with *s* as follows (Ref 7):

$$I \sim \exp(-10.25 \quad) \qquad \textbf{(Eq 1)}$$

where ϕ is the effective work function in eV (for most systems, this is on the order of a few eV) and *s* is in nm. One can see that if the tip-surface spacing is increased (decreased) by 0.1 nm, then the tunneling current will decrease (increase) by a factor of

$$\exp(1.025\sqrt{\phi}) \qquad \textbf{(Eq 2)}$$

or about a factor of 10 for an effective work function of 4 eV.

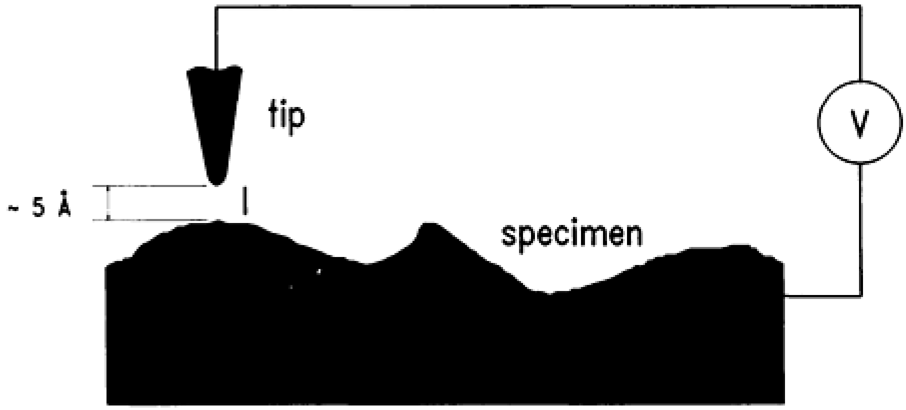


Fig. 1 Electron tunneling across gap between tip and specimen under an applied bias voltage

One can then exploit this sensitive dependence of *I* on the tip-surface spacing for topographic imaging, as described below. In scanning the tip horizontally across the specimen, any change in the tip-surface spacing results in a large change in *I*. One can use some feedback mechanism to move the tip up or down to maintain a constant value of *I*. According to Eq 1, this implies that one is maintaining a constant tip-surface spacing (the effective work function may change with position, because of surface heterogeneity, but its variation is ignored in this discussion). In other words, the up-and-down motion of the tip traces out the topography of the surface, analogous to the conventional technique of stylus profilometry, except that the tip never touches the surface in STM. This is known as constant current imaging, the most common imaging mode used in scanning tunneling microscopy.

Now the discussion returns to the situation where the effective work function varies with position. In the constant current imaging mode, one maintains a constant value during scanning. Therefore, for an absolutely flat surface with an effective work function of (4 ± 1) eV (a substantial surface heterogeneity) and a nominal tip-surface spacing of 0.5 nm, the STM topograph would reveal an apparent height variation of less than 0.1 nm, because of such surface heterogeneities. Hence, for surface roughness variations that exceed 0.1 nm, STM images obtained in the constant current mode reveal primarily surface topography.

To separately map such surface heterogeneities, one can modulate the vertical position of the tip in sinusoidal manner, for example, at a rate faster than the feedback response of the system. In this case, the tunneling current is also modulated. From Eq 1, one can readily show that:

$$\frac{d(\ln i)}{ds} \sim -\sqrt{\phi} \qquad \textbf{(Eq 3)}$$

from which the work function variation can be imaged.

Because of the proximity of the tip to the surface and the nature of tunneling, the tunneling electron beam diameter can be very small, and is given by (Ref 8):

$$\sqrt{0.2\,(R+s)} \qquad \textbf{(Eq 4)}$$

for an effective work function of 4 eV, where *R* is the radius of the tip, and *s* is the tip-surface spacing, both in units of nm. Therefore, for *R* = 0.2 nm and *s* = 0.5 nm, the electron beam diameter is on the order of 0.4 nm.

It is important to note that the STM works in vacuum, as well as in air and under liquids (Ref 9, 10). The reason is that the tip is so close to the surface (0.5 nm) that the volume through which the tunneling current flows is on the order of 0.1 nm^3. In a normal air environment, there is less than 0.003 air molecule contained in this volume. In water, the corresponding number is about 3. Therefore, the tunneling electrons flow more or less unaffected by ambient molecules.

Motion Control

To bring the tip to within tunneling range, it must be moved over macroscopic distances (hundreds of microns). During tunneling and feedback control, the tip must be positioned with precision better than 0.1 nm. The task is usually segregated into coarse and fine motion control.

Coarse motion control employs several methods. As described in their original work, Binnig and Rohrer used a piezoelectric inch-worm (piezoelectric materials are further described later). More recent designs are based on purely mechanical means, all using fine-thread screws (for example, Ref 11). Figure 2 depicts one type. Consider a conventional 80-pitch screw, that is, the screw advances by 25 mm (1 in.) after 80 turns. This translates into a motion advance of about 880 nm for 1° of screw rotation. Using a cantilever beam with a mechanical advantage of 10, as shown in Fig. 2, a precision of 88 nm (for 1° of screw rotation) can be achieved.

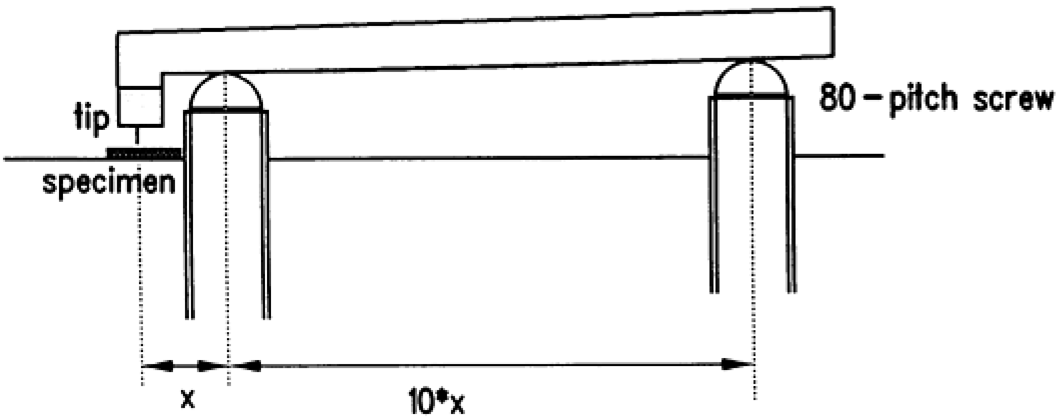


Fig. 2 Example of coarse approach using fine-thread screws and mechanical advantage of a cantilever beam

Fine Motion Control. During tunneling and feedback control, the precision required for tip positioning relative to the specimen surface must be better than 0.1 nm. This is achieved by using piezoelectric positioners. Piezoelectric materials expand or contract upon the application of an electric field (Fig. 3). Lead zirconium titanate (PZT) is the material of choice in the STM community. Most STMs are designed with a response of 1 to 10 nm/V, although a response of up to 500 nm/V can be achieved for large-scale scanning. Because voltages can be easily controlled and monitored in the submillivolt level, subnanometer control can be readily attained.

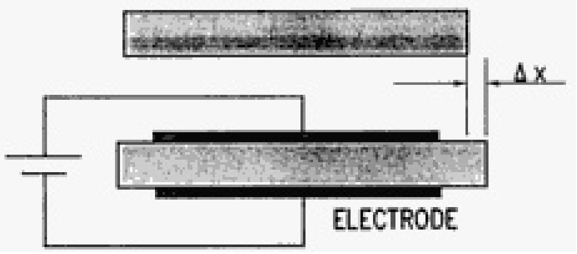


Fig. 3 Piezoelectric effect

In the initial development of STMs, three-axis control was accomplished using three separate pieces of piezoelectric bars held together in an orthogonal arrangement. To improve rigidity, especially for long-range scanners, most researchers opt for a design based on a piezoelectric tube scanner, shown in Fig. 4 (Ref 12). The inside of the piezoelectric tube is completely metal-coated, whereas the outside is metal-coated, in four separate quadrants. By applying appropriate voltages to one pair of diametrically opposite quadrants, one causes the piezoelectric tube to bend along that direction, thus achieving *x* or *y* scanning motion. Application of voltage to the inner surface causes the tube to expand or contract (*z*-axis motion). Therefore, three-axis motion can be attained with a single tube.

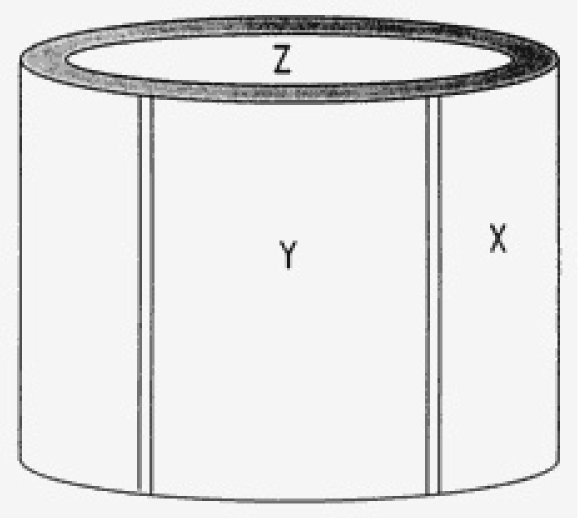


Fig. 4 Typical piezoelectric tube scanner used for STM work

The lowest resonance frequency of such a structure can be made to exceed 10 kHz easily (compared to, typically, 1 kHz for the orthogonal tripod). This higher resonance frequency allows electronic feedback and scanning to be performed at higher rates without setting the microscope into resonance or without crashing the tip onto the specimen surface. Therefore, a tube-based STM is inherently capable of acquiring topographical data rapidly. The major disadvantage of the tube scanner is the possibility of crosstalk among the three axes.

Vibration Isolation

The first tunneling microscope was supported using superconducting levitation for vibration isolation. More recent designs use damped springs, air tables, and stacked stainless steel plates separated by viton dampers. The goal of all these designs is to keep the tip-surface spacing immune to external vibration.

Assume that the STM sits on a platform that is coupled to the outside world via a spring with resonance frequency, *f*, and that the lowest resonance frequency of the STM is *F*, which is much greater than *f* (Fig. 5). The external vibration has a

frequency, f', and amplitude A. With such a system, the vibration amplitude transmitted to the STM depends on the frequency f' of the vibration. There are four regimes to consider.

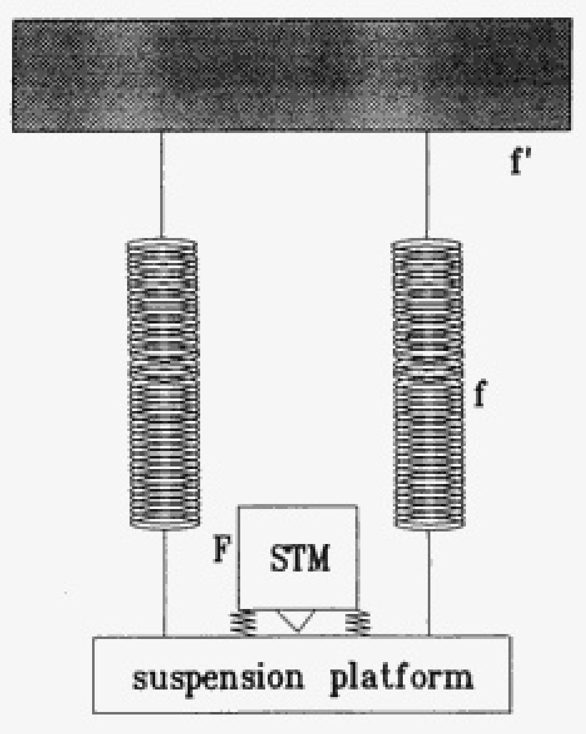


Fig. 5 Vibration isolation for an STM using a suspension platform. Source: Ref 11

The first regime is $f' < f$. The platform spring does nothing to attenuate the external vibration. The vibration amplitude entering into the microscope causes a tip-surface spacing change given by

$$a = A\,(f'/F)^2 \qquad \textbf{(Eq 5)}$$

The second regime is $f' \sim f$. The vibration amplitude entering into the microscope is actually amplified, depending on the amount of damping in the platform spring.

The third regime is $f \ll f' \ll F$, where the vibration amplitude, a, entering into the microscope is independent of f' and is given by

$$a = A\,(f/F)^2 \qquad \textbf{(Eq 6)}$$

In the fourth regime, $f' \gg F$, only the platform spring does the attenuation and the transmitted vibration amplitude, a, is given by

$$a = A\,(f/f')^2 \qquad \textbf{(Eq 7)}$$

In examining these four cases, it becomes clear that one should support the STM on a soft platform (small f value) and design a microscope with high rigidity (large F value). For example, for the third regime, with $f = 1$ Hz, $F = 10$ kHz, and $A = 10$ μm (400 μin.), the transmitted vibration amplitude, a, can be shown to be equal to 10^{-4} nm for intermediate frequencies. Therefore, a rigid STM not only allows fast image acquisition, but more effective vibration isolation, as well.

Tip Preparation

The two types of tip materials that are widely used are tungsten and platinum alloys (for example, Pt-Ir and Pt-Rh). Tungsten is strong, and it can be fabricated into a sharp tip easily, although it tends to oxidize rapidly in air. On the other hand, Pt alloys are stable in air, but they may not survive occasional tip crashes on surfaces.

Several methods can be used to create sharp tips of these materials. These include electropolishing (Ref 13, 14), cutting and grinding, momentary application of a high-bias voltage (a few volts), or simply waiting for a few minutes after setting up in the tunneling configuration.

Recently, a simple technique was described by Akama and coworkers, in which a sharp, amorphous carbon needle was grown by aiming a focused 25 keV electron beam at the apex of a tungsten tip inside a scanning electron microscope (SEM) for a few minutes (Ref 15). Decomposition of ambient hydrocarbons results in the formation of a carbon needle at the apex. A carbon tip several microns long with an apex radius less than 25 nm can be produced readily. Tips with such an elongated geometry are useful to image rough surfaces with minimal artifacts.

Data Acquisition and Analysis

A typical setup is shown in Fig. 6 (Ref 11). A voltage bias is applied between the tip and the specimen. The tunneling current so obtained is then compared with a preset value between, typically, 1 and 10 nA. The error signal then drives an integral feedback circuit, the output of which is used to control a fast high-voltage operational amplifier, which feeds voltage to the *z* electrode of the tube scanner. At the same time, raster-scanning is accomplished by using two digital-to-analog converters to control the output of two high-voltage operational amplifiers feeding voltages to the *x* and *y* electrodes of the tube scanner.

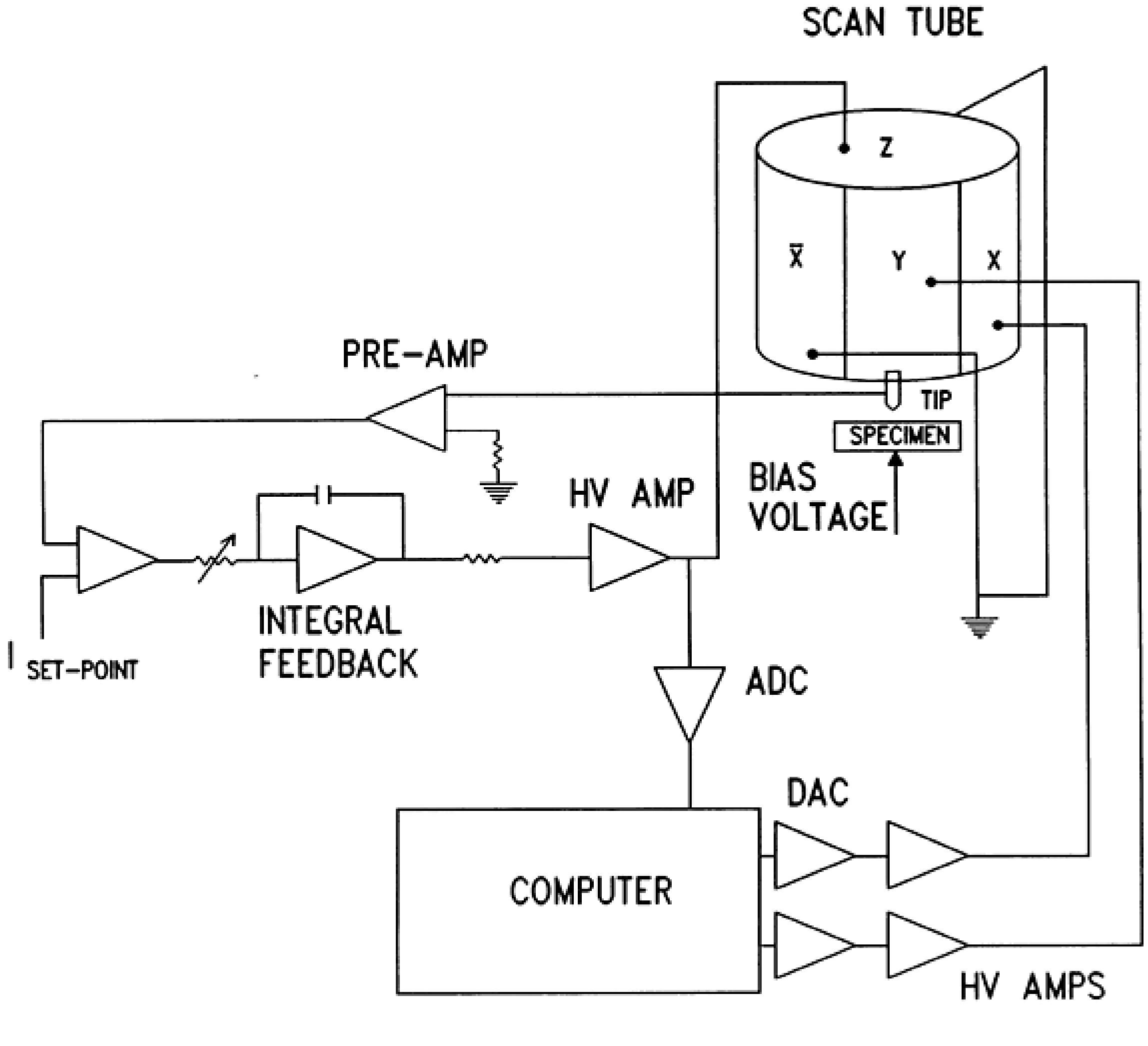


Fig. 6 Data acquisition and control electronics for STM

At each step, the *z* voltage required to maintain a constant tunneling current is read by the computer via a fast analog-to-digital converter (either through AC coupling or potential dividers). This *z* voltage, as discussed earlier, corresponds to the surface height at that *xy* location. This information can then be displayed in real time as gray-level images on an analog monitor. Most tube-based scanners allow image acquisition at the rate of several thousand pixels per second.

In the above setup, the feedback control is done by analog circuitry. Because of the availability of fast digital signal processors in recent years, microscopes can be designed to use signal processors to perform feedback via software.

Normally, the topographical data are stored a two-dimensional integer arrays. As a result, each image can be processed in a variety of ways, for example, to suppress noise, to enhance parts of the image, or to obtain certain surface roughness parameters.

Several commercial software packages are available for these types of image processing and analysis setups on personal computers and workstations. Hard-copy outputs can be obtained as either line plots or gray-level images. In the latter case, several methods are available. The most direct method is to photograph the monitor screen or use a film recorder to obtain an image. Another convenient method is to use a laser printer, color wax printer, or video printer.

Applications

A standard benchmark in testing STM performance is to obtain an atomic image from highly oriented pyrolytic graphite (HOPG) in air. Figure 7 shows an image obtained from HOPG in air, measuring about 1.4 nm × 1.4 nm. Bright protrusions are individual carbon atoms on the basal plane of graphite. The height of the protrusions is about 0.2 nm.

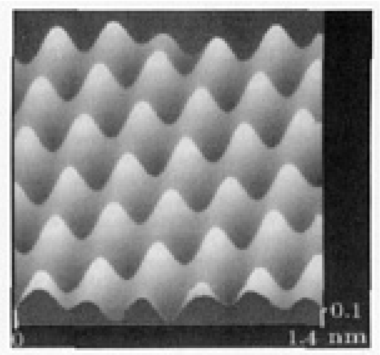


Fig. 7 Atomic image of HOPG graphite obtained in air

Figure 8 shows a 2.5 × 2.5 μm (100 × 100 μin.) region of a silver specimen that has been fatigued in oxygen (Ref 16). The stress axis is parallel to the *y* direction in the image. The image shows three parallel slip bands. Figure 9 shows a large-area scan from a silicon substrate with aluminum lines deposited on it.

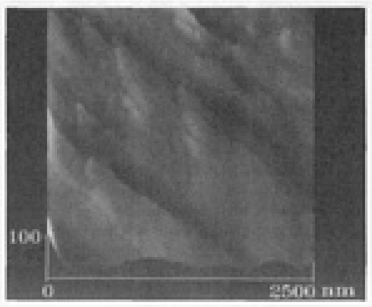


Fig. 8 Topography image from a fatigued silver single crystal showing parallel slip bands; faint vertical lines are due to polishing

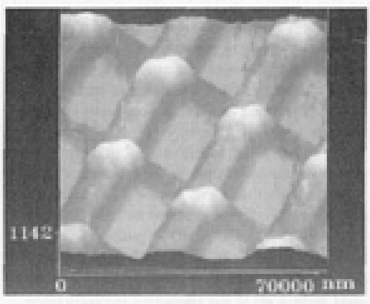


Fig. 9 Topography image of a large (70 × 70 μm, or 2.8 × 2.8 mil) region of a silicon wafer with aluminum lines. Specimen was coated with 50 A (2 × 10^{-7} in.) of gold to improve conductivity.

Scanning tunneling microscopy is normally applicable only to conducting surfaces because a conducting path is needed for the tunneling electrons. An interesting situation arises when a conducting surface is covered by some insulating species. For example, Koch and Hamers (Ref 17) found that when tunneling electrons of the proper energy are injected onto thin oxide patches on an otherwise clean silicon surface, the oxide traps the electrons and then releases them after some time. As a result, the tunneling current exhibits large amplitude fluctuations. On the other hand, no such large amplitude fluctuations are observed when the current is injected onto the clean surface.

Consider a map of tunneling current fluctuations as a current variance image. The image is obtained by measuring the fluctuation amplitude of the tunneling current at each point of the surface. The fluctuation amplitudes are then displayed as gray-level maps, with brighter regions indicating greater fluctuations.

One interesting application (Ref 18) is to image the distribution of thin layers of fluorocarbon lubricants applied on magnetic thin-film disk surfaces (fluorocarbon is an insulator). Figure 10 shows topography and variance images from scratched and unscratched portions of a magnetic thin-film disk. The topography images were obtained at a low bias (20 mV), whereas the variance images were obtained at a high bias (4 V). The variance image from the unscratched surface shows uniformly high variance, suggesting uniform distribution of the lubricant. The variance image from the scratched surface shows regions with low current fluctuations. This indicates that the scratched regions have bare spots.

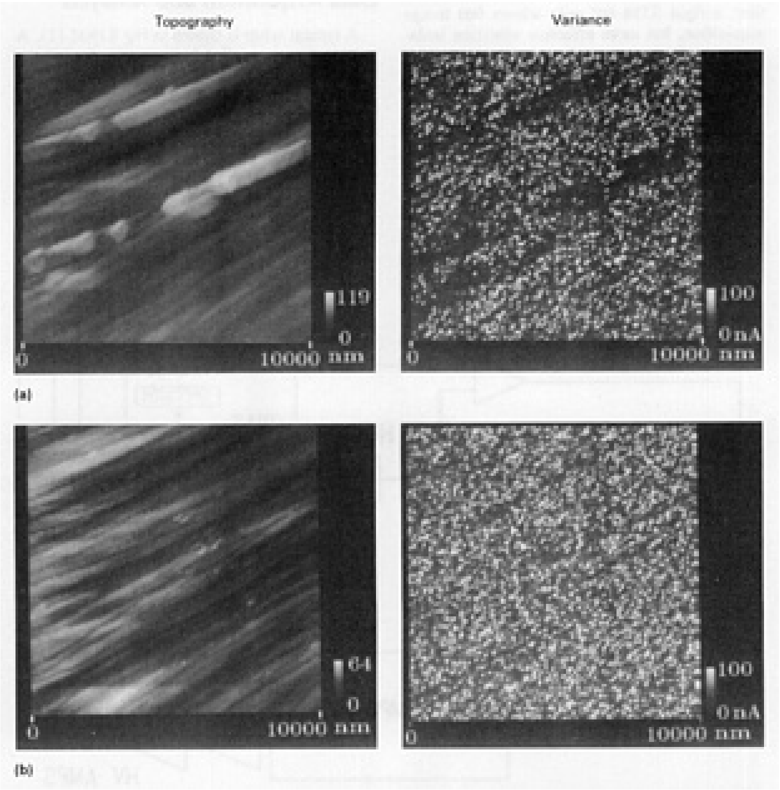


Fig. 10 Topography and variance images from scratched (a) and unscratched (b) regions of a magnetic thin-film rigid disk coupon. Source: Ref 18

Technique Comparison

Table 1 summarizes the characteristics of several other surface roughness measurement techniques, specifically, stylus profilometry, optical interferometry, and SEM stereomicroscopy (Ref 11). Although these techniques are easy to use, they have limitations.

Table 1 Comparison of several surface roughness measurement techniques

Type	Contact	Resolution, nm	
		Lateral	**Vertical**
Stylus profilometry	Yes	1000	5
Optical interferometry	No	500	1
SEM stereo-microscopy	No	10	60

For example, the lateral resolution in stylus profilometry is limited by the radius of the diamond tip, typically a few microns. Surface features with dimensions less than the tip radius are not resolved. There is also the possibility of specimen damage induced by the stylus during measurements. In principle, these problems can be overcome by using a sharp stylus and an ultralight load. Optical interferometry overcomes most of these problems, but is diffraction-limited in lateral resolution to a fraction of a micron (unless near-field imaging techniques are used). Stereomicroscopy in which an SEM is used gives relatively poor vertical resolution and requires placing the specimen in vacuum. In contrast, the STM achieves superb horizontal and vertical resolution and can be applied in virtually any environment.

Limitations and Solutions

The STM has two major limitations. First, the specimen surface must be reasonably conducting. Under typical operating conditions, the resistance of the gap separating the tip and the specimen is on the order of 10 MΩ (for example, tunneling

at 1 nA under a bias of 10 mV). "Reasonably conducting" means that the resistance of the electrical path from the specimen to the return circuit should be small, compared to 10 MΩ. This rules out the consideration of many ceramic and polymer materials. One solution is to put a conduction coating (such as gold or Pt-C) on such surfaces, assuming that the coating faithfully reproduces the surface topography of the substrate. Another solution is to use AC tunneling, that is, the bias is allowed to change sign rapidly (Ref 19).

The basic idea is that in the forward cycle, electrons are injected from the tip onto the surface. The behavior of the tunneling current with respect to tip-surface spacing is as predicted by Eq 1 in this portion of the cycle. Therefore, feedback control can be "locked" to the tunneling current in the forward cycle. In the next half cycle, the polarity is reversed, thereby clearing electrons from the surface of the insulating specimen. In this way, insulating surfaces can be imaged by the STM. The major difficulty is that high-frequency AC bias may be required for highly insulating surfaces. Stray capacitance between the tip and the sample may result in a large displacement current that can overwhelm the tunneling current.

Second, the STM suffers from limited scanning range. Using reasonable geometry (for example, a scan head on the order of a few centimeters long and a scan tube thickness of 1 to 2 mm, or 0.04 to 0.08 in.) and applied voltage (not exceeding 300 V), one finds that the maximum scan range is on the order of from 50 to 100 μm (2 to 4 mils). In general, because longer scanners have lower resonance frequencies, scanning rates must be reduced to obtain images over large areas. The important point is that the STM cannot replace the conventional SEM in terms of dynamic range.

Future Trends

In spite of its short history, the STM has developed into a mature tool for a wide range of applications in many science and engineering disciplines. For tribologists who are interested in surface roughness measurements, the new technique offers high resolution in both horizontal and vertical directions with few environmental constraints (in terms of vibration isolation and measurements under normal ambient conditions). Of course, tips with large aspect ratios are still required to image rough surfaces with minimal artifacts.

There are two important variants of the STM: the scanning capacitance microscope (SCaM) and the atomic force microscope (AFM). These new microscopy techniques allow the imaging of surface topography, and obtain other important surface properties, as well.

In a SCaM, the capacitance between the tip and the specimen surface is used as a sensor of the tip-surface spacing (Ref 20). Despite the very small capacitance involved in these measurements, spatial resolution of about 25 nm has been demonstrated. In addition, by exploiting the fact that the capacitance of a semiconductor surface depends on the carrier concentration, this technique can be used to image dopant distribution on semiconductor surfaces at high spatial resolution (Ref 21).

In an AFM, the tip is normally part of a small wire (Ref 22, 23) or a microfabricated cantilever (Ref 24). The force of interaction between the tip and the surface results in a deflection of the cantilever. In most modern designs, the cantilever deflection is sensed either by detecting the reflection of a light beam from the back of the cantilever (Ref 25) or by optical interferometry (Ref 23, 24). Under appropriate operating conditions, atomic resolution can be achieved (Ref 24). One important strength of the AFM is its ability to obtain images from insulator surfaces. By using a magnetized tip with the AFM, magnetic domains have been imaged (Ref 26).

Extensions of STM, such as the AFM, will continue to evolve. For example, by adapting the AFM to measure forces in the vertical and horizontal direction, one can turn it into a tribometer measuring frictional forces on the atomic scale (Ref 27).

In parallel with developments in research laboratories around the world, commercial versions of STM have been developed for applications in air and ultrahigh vacuum (Ref 28). Associated turnkey image processing software packages are widely available for different types of computers, thus making data analysis and presentation extremely convenient. With these developments, the STM should eventually become standard equipment in most analytical laboratories.

References

1. R. Young, Surface Microtopography, *Phys. Today,* Vol 24 (No. 11), 1971, p 42

2. R. Young, J. Ward, and F. Scire, The Topografiner: An Instrument for Measuring Surface Microtopography, *Rev. Sci. Instrum.,* Vol 43, 1972, p 999
3. G. Binnig, H. Rohrer, C. Gerber, and E. Weibel, 7 × 7 Reconstruction on Si(111) Resolved in Real Space, *Phys. Rev. Lett.,* Vol 50, 1983, p 120
4. M.A. McCord and R.F.W. Pease, Lithography with the Scanning Tunneling Microscope, *J. Vac. Sci. Technol. B,* Vol 4 (No. 1), 1986, p 86
5. E.E. Ehrichs, R.M. Silver, and A.L. de Lozanne, Direct Writing with the Scanning Tunneling Microscope, *J. Vac. Sci. Technol. A,* Vol 6 (No. 2), 1988, p 540
6. M. Ringger, H.R. Hidber, R. Schlogl, P. Olehafen, and H.-J. Guntherodt, Nanometer Lithography with the Scanning Tunneling Microscope, *Appl. Phys. Lett.,* Vol 46 (No. 9), 1985, p 832
7. P.K. Hansma and J. Tersoff, Scanning Tunneling Microscopy, *J. Appl. Phys.,* Vol 61 (No. 2), 1987, p Rl
8. J. Tersoff and D.R. Hamann, Theory of the Scanning Tunneling Microscope, *Phys. Rev. B,* Vol 31 (No. 2), 1985, p 805
9. R. Sonnenfeld and P.K. Hansma, Atomic Resolution Microscopy in Water, *Science,* Vol 232, 1986, p 211
10. S.M. Lindsay and B. Barris, Imaging Deoxyribose Nucleic Acid Molecules on a Metal Surface Under Water by Scanning Tunneling Microscopy, *J. Vac. Sci. Technol. A,* Vol 6 (No. 2), 1988, p 544
11. T.-S. Lin, Y.-W. Chung, and H.S. Cheng, High Resolution Surface Roughness Measurements in Air Using a Scanning Tunneling Microscope, *Lubr. Eng.,* Vol 46 (No. 5), p 304
12. G. Binnig and D.P.E. Smith, Single-Tube Three-Dimensional Scanner for Scanning Tunneling Microscopy, *Rev. Sci. Instrum.,* Vol 57, 1986, p 1688
13. A.A. Gewirth, D.H. Craston, and A.J. Bard, Fabrication and Characterization of Microtips for In Situ Scanning Tunneling Microscopy, *J. Electroanal. Chem.,* Vol 261, 1989, p 477
14. G.A. Wardly, "A Procedure for Fabricating Tungsten Field Emission Tips and Related Phenomena," Report RC 4455, IBM, 30 July 1973
15. Y. Akama, E. Nishimura, and A. Sakai, New Scanning Tunneling Microscopy Tip for Measuring Surface Topography, *J. Vac. Sci. Technol. A,* Vol 8 (No. 1), 1990, p 249
16. T.S. Sriram, M.E. Fine, Y.-W. Chung, STM and Surface Analytical Study of Fatigue Crack Initiation in Silver Single Crystals: Effects of Oxygen Partial Pressure, *Scr. Metall.,* Vol 24, 1990, p 279
17. R.H. Koch and R.H. Hamers, Characterization of Electron Trapping Defects on Silicon by Scanning Tunneling Microscopy, *Surf. Sci.,* Vol 181, 1987, p 333
18. T.S. Sriram, K.J. Wahl, Y.-W. Chung, B. Bhushan, and W. Rothschild, The Application of Scanning Tunneling Microscopy to Study Lubricant Distribution on Magnetic Thin-Film Rigid Disk Surfaces, *J. Tribol.,* Vol 113, 1991, p 245
19. G.P. Kochanski, Nonlinear Alternating Current Tunneling Microscopy, *Phys. Rev. Lett.,* Vol 62 (No. 19), 1989, p 2285
20. J.R. Matey and J. Blanc, Scanning Capacitance Microscopy, *J. Appl. Phys.,* Vol 47, 1985, p 1437
21. C.C. Williams, W.P. Gough, and S.A. Rishton, Scanning Capacitance Microscopy on a 25 nm Scale, *Appl. Phys. Lett.,* Vol 55 (No. 2), 1989, p 203
22. G. Binnig, C.F. Quate, and C. Gerber, Atomic Force Microscope, *Phys. Rev. Lett.,* Vol 56 (No. 9), 1986, p 930
23. R. Erlandsson, G.M. McClelland, C.M. Mate, and S. Chiang, Atomic Force Microscopy Using Optical Interferometry, *J. Vac. Sci. Technol. A,* Vol 6 (No. 2), 1988, p 266
24. T.R. Albrecht and C.F. Quate, Atomic Resolution with the AFM on Conductors and Nonconductors, *J. Vac. Sci. Technol.,* Vol 6 (No. 2), 1988, p 271
25. Y. Martin, C.C. Williams, and H. Wickramasinghe, Atomic Force Microscope: Force Mapping and Profiling on a Sub-100 A Scale, *J. Appl. Phys.,* Vol 61, 1987, p 4723
26. J.J. Saenz, N. Garcia, P. Grutter, E. Meyer, H. Heizelmann, R. Wiesendanger, L. Rosenthler, H.R. Hidber, and H.-J. Guntherodt, Observation of Magnetic Forces by the Atomic Force Microscope, *J. Appl. Phys.,* Vol 62, 1987, p 4293

27. C.M. Mate, G.M. McClelland, R. Erlandsson, and S. Chiang, Atomic Scale Friction of Tungsten on a Graphite Surface, *Phys. Rev. Lett.*, Vol 59, 1987, p 1942

28. *Physics Today Buyers Guide,* August 1990

Measurement of Surface Forces and Adhesion

Roger G. Horn, Ceramics Division, National Institute of Standards and Technology

Introduction

BECAUSE ADHESION between two contacting surfaces is an important factor in terms of friction and wear, the ability to make controlled, reproducible measurements of this quantity is desirable. The methods to accomplish this and some of the significant difficulties involved in doing so are described in this article.

The common sense of the word "adhesion," that is, the joining of two materials with a thin film of a third material, is not of concern here. The engineering and chemical aspects of adhesives and adhesive joints are described in Ref 1 and 2. Rather, the concern is the innate adhesion between two materials that arises from interatomic and intermolecular forces. Such forces are always present, and act between all of the atoms and molecules in the system--not only within each material, but also those in different materials and those in any other media present. For example, two solid surfaces in close proximity are attracted as a result of all the van der Waals forces acting between the atoms of each solid. Furthermore, the interaction is affected by the molecules of whatever liquid or gas separates the solids.

Intermolecular forces that act between two separate bodies can generally be considered as acting between the surfaces of those bodies, and are therefore called surface forces (Ref 3, 4). Adhesion is one manifestation of the existence of surface forces. This article first describes surface forces and methods of measuring them before adhesion is discussed. Surface force measurements help one to understand the importance of various contributions to surface forces and adhesion, and therefore, ultimately, to design materials and environmental conditions to minimize or maximize adhesion, as required.

Basic Concepts

Surface forces have their origins in well-understood interatomic and intermolecular forces (Ref 3), but several indirect effects can occur to enrich the variety of surface forces that is possible in a given situation (Ref 4). It is convenient to classify surface forces according to their range of action. Short-range forces are those that act between atoms and molecules that are essentially in contact, say within 0.1 or 0.2 nm of each other. Examples are covalent and hydrogen bonding, as well as Born repulsions. Long-range forces act between surfaces that are further apart; long range in this context might mean a few nanometers. Examples are van der Waals and electrostatic forces.

To complicate this classification, forces that are effective at long range, but that are really the result of short-range forces, also exist. An example is steric repulsion, in which surfactant molecules adsorbed on a solid surface (via short-range bonding that can be van der Waals, ionic, covalent or hydrogen bonding) prevent a second surface from a approaching the first (via Born repulsions, either between the adsorbed molecules and the second surface of between one layer of adsorbed molecules and another layer). Boundary lubricants operate in this way to keep solid surfaces separated.

The study of forces between molecules and surfaces does not fall within the compass of "surfaces forces" in the sense used here, but they can be very important for the reason described above. Measurement of forces between one surface and another can often give indirect information on molecule-surface forces. For example, the presence or absence of an adsorbed layer can easily be detected using the methods described in this article.

In general, short-range forces are stronger than long-range forces, and make the most important contributions to adhesion. Unfortunately, long-range forces are easier to both measure and model. Therefore, knowledge of surface forces, both theoretical and experimental, is much sounder for long-range effects, and adhesion remains an area of difficulty.

Surface energy is the work required to create a unit area of new surface from bulk material. A though experiment is commonly employed to define surface energy, γ_s, as half of the reversible work, W_c, required to first divide a monolithic solid so as to create two new surfaces of unit area, and then to pull the two surfaces infinitely far apart in vacuum. Here, infinitely far apart means beyond the range of any surface force; usually, a few μm is ample. The term W_c is called the work of cohesion. In general, the separation process involves breaking interatomic bonds, the number (and possibly strength) of which depends on crystallographic orientation, so that surface energy varies from one crystal face to another (Ref 5).

Surface Energy and Surface Forces. In conducting the thought experiment described above, the act of separating two surfaces would require work to be done against the attractive surface forces present. Integration of this surface force, as a function of separation from 0 to ∞, gives a quantity

$$W_{SF} = -\int_0^\infty F_f(D)\, dD \qquad \textbf{(Eq 1)}$$

where $F_f(D)$ is the surface force between flat solids at a separation, D, using the convention that $F < 0$ for an attractive force. The subscript SF indicates that this is the contribution to W_c from the surface forces.

Sometimes W_{SF} is equated to twice the surface energy, but this is not generally correct because other energetic processes are likely to occur during the creation of new surfaces. In particular, the new surfaces are likely to reconstruct. Atoms near the new surface rearrange their positions to a configuration that is more favorable near a surface than was their original configuration in the bulk of the solid. Furthermore, there is the delicate issue of assigning the energy associated with breaking strong, short-ranged atomic bonds when the original solid is "magically" cleaved. This is certainly a contribution (probably the major one) to W_c, but it is unclear whether or not it should be included in the definition of W_{SF}.

Interfacial Energy. When a solid is not in vacuum but is in contact with a liquid or vapor, it is considered in terms of interfacial energy, γ_{SL} or γ_{SV}, respectively, rather than surface energy. If the above process were carried out in either a liquid or vapor environment, the work of cohesion would equal twice the interfacial (solid-liquid or solid-vapor) energy. The same remarks about reversibility and reconstruction still apply. Furthermore, in this situation there is also the likelihood that molecules of the liquid or vapor will adsorb to the new solid surface (which could be thought of as "reconstruction" of the fluid at the interface). Adsorption reduces the interfacial energy.

Note that the reconstruction of both solid and fluid depends on which materials are involved. For example, a given solid material may reconstruct differently in water than in a hydrocarbon liquid. The details of the interfacial structure are also likely to depend on the distance between one interface and the other. Again, careful consideration must be given in the thought experiment to the question of reversibility. If the adsorption is irreversible, then bringing the two solids back together after they have been separated will not remove the adsorbate, and the surfaces will never return to their original intimate contact and will not reform the original atomic bonding that existed in the monolithic solid.

The interfacial energy, γ_{LV}, between a liquid and its own vapor is also called the surface energy, or, more commonly, the surface tension of that liquid. The term "tension" comes from a real force that resists any attempt to increase the surface area of a given volume of liquid. Surface tension of a liquid can be measured directly. Determining either the surface or interfacial energy of a solid is much more problematic. Solid-liquid and solid-vapor surface energies are related to the liquid surface tension and the contact angle, θ, made by the liquid on the solid through Young's equation:

$$\gamma_{SV} = \gamma_{SL} + \gamma_{LV} \cos\theta \qquad \textbf{(Eq 2)}$$

Work of Adhesion. A thought experiment has been used to define the work of cohesion of a solid as twice the surface energy (because *two* surfaces are created) or interfacial energy, as appropriate. If a similar experiment were done in which

two *different* materials were separated from contact to a large distance, the work required would be the work of adhesion, W_A.

The work of adhesion can be equated with the energetic cost of creating new interfaces of each solid material in contact with the environment, γ_{1E} and γ_{2E}, say, where the subscript $_E$ could be liquid or vapor, as appropriate, less the energetic gain from removing the original solid-solid interface. The energy associated with the solid-solid interface is also termed an interfacial energy, γ_{12}, which is defined by the Dupré equation:

$$W_A = \gamma_{1E} + \gamma_{2E} - \gamma_{12} \quad \textbf{(Eq 3)}$$

In the discussion below, adhesion will sometimes be loosely used to mean work of adhesion.

In a similar way, one could define the work of adhesion between two pieces of the *same* solid material. This would differ from the work of cohesion if the two pieces were brought together in an environment so that adsorbate remained at the interface or if they were formed with their crystal orientations misaligned, giving a grain boundary at the interface. In that case, the quantity γ_{12} would be nonzero, and would correspond to a grain boundary energy, with or without extrinsic molecules present. Because grain boundary energy depends on grain boundary angle, the adhesion between two crystals of the same material depends on their relative orientation.

Note that there is another effect that can make the above definition of work of adhesion problematic. When two solids are pulled apart, it is quite possible, and often observed in practice, that separation does not occur precisely between the two materials. Many solids interdiffuse when in contact, particularly at high temperature, which makes it almost impossible to separate them in the ideal way that the simple thought experiment would imply. Even without interdiffusion, it is frequently found that joints break parallel to, but not precisely at, the interface. Many experimentalists have detected at least one of the materials on *both* sides of the division following separation (Ref 5).

When the environment includes a condensable vapor, an additional contribution to the adhesion between two solids occurs from an effect known as capillary condensation. Liquid condenses (as long as the contact angle is less than 90°) to form a bridge wherever the gap between the solids is small. Curvature of the liquid-vapor meniscus results in a negative Laplace pressure in the liquid, which acts as a cohesive force that holds the solids together. The magnitude of this force depends on the relative vapor pressure of the liquid, and on the geometry of the solids (Ref 3, 6). In some situations, this is the predominant factor in adhesion.

Measuring Surface Forces

Instrumental Requirements. This section discusses measurement of the surface forces previously described as "long range." However, the forces are weak and very short-ranged when considered on a laboratory scale. Typical forces range from 10^{-7} to 10^{-4} N, acting at surface separations from 1 to 100 nm. Therefore, measuring them is not a trivial matter. The magnitude of the forces increases with surface area (or, as shown below, with the radius of curved surfaces). Thus, in order to have measurable forces, it is desirable to have extended areas of surface. At the same time, to make sensible measurements of surface separation on a nanometer scale, it is necessary for the surfaces to be extremely smooth. A method of detecting small forces is required, as are methods of controlling and measuring very small surface separations.

Techniques. Although surface force measurements have been made for several decades, the inherent difficulties of surface preparation and cleanliness limited the number of materials studied and the amount of data to a small level, until about 1970. A review of the field up until 1982 is provided in Ref 7. Currently, the most popular technique is the crossed-cylinder apparatus devised by Tabor (Ref 8) and further developed by Israelachvili (Ref 9). This technique uses a surface force apparatus (SFA), which is commercially available.

The SFA consists of a closed stainless steel chamber designed to enclose a variety of liquid or vapor media, and is usually operated ambient temperature and pressure. Force is measured between two cylindrical surfaces, with the axes of the cylinders at right angles to each other (Fig. 1). The reason for choosing this geometry is one of experimental practically. If two planar surfaces were to be used (as one might have supposed), then there would be extreme difficulties in forming surfaces of sufficient flatness, mounting them exactly parallel, maintaining parallelism while moving the surfaces together or apart, and avoiding edge effects, all while working on a nanometer scale. As shown in the next section, "The Derjaguin Approximation," it turns out that there is no difficulty in interpreting the forces measured in this odd geometry, because

there is a simple relationship between the force measured between crossed cylinders and the energy of interaction between flat surfaces.

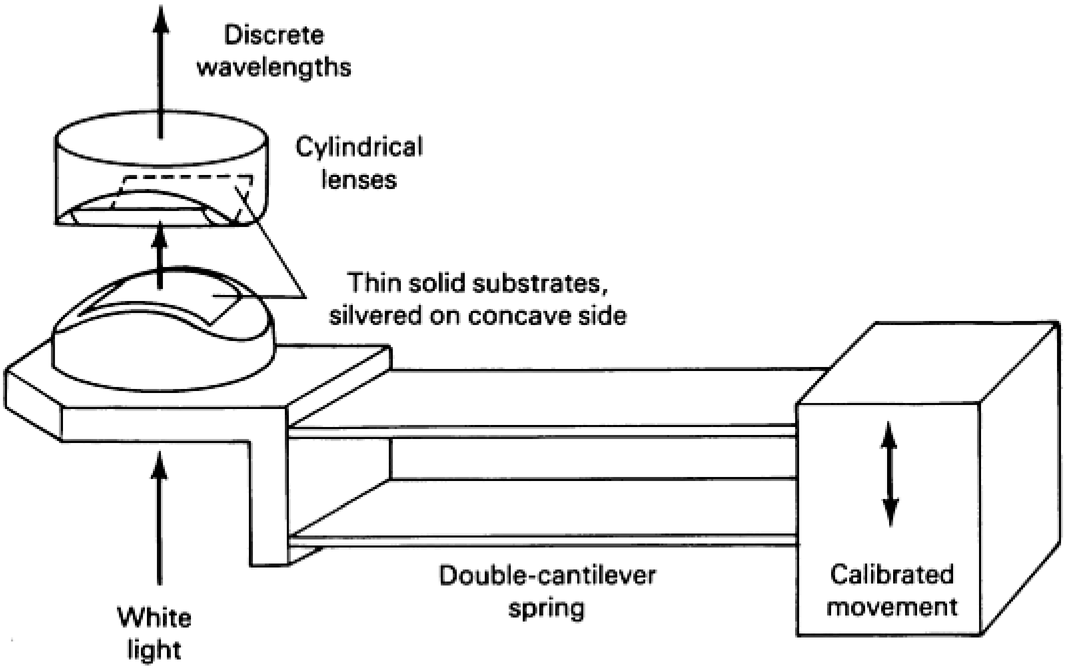


Fig. 1 Surface force apparatus, in which two thin solid substrates are mounted as crossed cylinders, with one of them supported by a cantilever spring whose deflection measures the force. An optical interferometric technique is used to measure the distance between the surfaces.

In order to measure surface separation, the SFA employs an optical interference technique (Ref 9, 10, 11). Under optimal conditions, this gives a resolution of 0.1 nm or better. Of course, one drawback is that it places a limitation on the solid materials that can be investigated, namely, that at least one of the pair whose surfaces approach contact must be transparent and rather thin (ideally, a few μm). As discussed further in the section "The Pull-Off Force," most of the measurements made with this apparatus have been made with thin foils of mica bent around and glued to cylindrical glass lenses.

To implement the optical method, a 95% reflecting silver layer is coated on the outer (that is, remote) surface of each solid substrate. Collimated white light is shone through the two substrates and whatever medium separates them (Fig. 1). Multiple-beam interference between the two silver layers selects only certain wavelengths of light, which are passed by the interferometer. All other wavelengths interfere destructively and are not transmitted. The transmitted light is collected and directed to a grating spectrometer, which spreads it according to wavelength, so that discrete wavelengths appear at the exit port of the spectrometer as spatially separated fringes of equal chromatic order.

The wavelengths depend on the thicknesses and refractive indices of the materials that are included in the interferometer: usually, the two transparent substrates and whatever fluid medium is between them. Measurement and analysis of the wavelengths allow computation of these thicknesses (Ref 10, 11). Because the two solids are of fixed thickness, those values can be subtracted from the total to give the thickness of the intervening medium, that is, the separation, D between the inner (adjacent) surfaces of the solids at their closest point of approach.

The zero value of separation is defined in practice as the position that the two solids come to, under the influence of an adhesive force (and without external load). This is a sound operational definition of $D = 0$, but it begs the question of precisely where the solid surfaces are with respect to each other. In many situations (for example, surfaces brought into contact in air), it is probable that one or more layers of adsorbate molecules remain trapped between the surfaces.

The interferometer must have a minimum thickness in order to give transmitted wavelengths in the visible region of the spectrum, but should also be rather thin to give a high resolution in measuring thickness changes. In practice, optimal

resolution is obtained when the total thickness (between the silver layers) ranges from about 3 to 10 μm (120 to 400 μ in.).

The surface force that one solid substrate exerts on the others is measured by a simple spring-deflection method. One solid is mounted on a cantilever spring, the remote end of which is moved up or down using a three-stage drive mechanism. The first stage is a micrometer that allows coarse positioning of the surfaces from a separation of a few mm to a few μm. The second stage is a micrometer that acts through a differential spring mechanism, which reduces the motion a thousand-fold, allowing positioning to approximately 1 nm. Finally, voltage applied to a piezoelectric tube expander gives positioning to a fraction of 1 nm.

After calibrating the drive mechanisms, it is straightforward to monitor any differences between a movement of the remote end of the spring and the distance moved by the end that bears one of the solids. This difference corresponds to a deflection of the spring. Multiplying it by the spring stiffness (typically 100 N/m, or 7 lbf/ft) gives the increment in force resulting from the movement. Because both the calibration and the movement of the solid are measured with a resolution of ~0.1 nm, it is possible to measure very small force changes (10^{-7} to 10^{-8} N) using this technique.

The Derjaguin Approximation. The force, F_c, between two gently curved surfaces is proportional to the interaction energy per unit area, E_f, between two flat ones at the same separation. This relationship, known as the Derjaguin approximation, allows straightforward interpretations to be made of surface force measurements between crossed cylinders (or between one sphere and another or between a sphere and a flat plate). It is also helpful in certain adhesion measurements, as described below.

The Derjaguin approximation is derived (see Ref 3, for example) by considering the force between each element of one curved surface and each element of the other, and then integrating over the two surfaces to obtain the total force. As long as the radius of curvature is much larger than the range of the surface force, this is approximately equivalent to integrating the force per unit area, F_f, between flat surfaces, from the minimum separation of the curved surfaces, D, to an effectively infinite upper limit, with some geometrical factors to account for the shape of the surfaces. The integral simply gives the interaction energy between flats, $E_f(D)$, which is the work done against the surface forces in moving the flat surfaces from infinity to D. For two spheres of radius R_1 and R_2, the geometrical factor is a constant, giving

$$F_c(D) = 2\pi R E_f(D) \quad \textbf{(Eq 4)}$$

where $1/R = 1/R_1 + 1/R_2$. It can be shown that the geometry of crossed cylinders of equal radii, R_c, is equivalent to a sphere of radius R_c approaching a flat plate, or to two spheres of radius $2R_c$ approaching each other.

The Pull-Off Force. A special case of the Derjaguin approximation occurs when two rigid curved bodies are in contact at a point, that is, $D = 0$. In this situation, the force between them is $2\pi R E_f(0) = -2\pi R W_{SF}$, because (from Eq 1), $E_f(0)$ is just $-W_{SF}$. This is also the magnitude of the minimum force required to separate the sphere from the plate, which is called pull-off force, F_p. (The pull-off force has the opposite sign as that of the surface force: $F_p > 0$ when there is attraction.)

Substrate Materials. As indicated above, the original and still most common solid material used in surface force measurements is mica, chosen because it satisfies the requirements (thin and transparent) of the optical interference technique used in the SFA and because it is easy to prepare large areas of molecularly smooth surface by cleavage. Experiments have been conducted on mica surfaces immersed in many different liquid and vapor environments (Ref 4).

Recently, there has been some success in extending these measurements to a wider range of surfaces. One approach is to coat mica surfaces by various techniques, including Langmuir-Blodgett deposition, surfactant or polymer adsorption from solution, plasma modification, and evaporative coating of thin metal, carbon, and metal-oxide films. An alternative approach is to find a means of preparing other transparent materials as micron-thick foils with very smooth surfaces. This has now been done for sapphire, silica, pyrex glass, and certain polymers. It is reasonable to expect that the range of materials studied will continue to increase in the near future.

Currently, the best way to prepare metal surfaces for the SFA appears to be thin-film evaporation onto mica or another smooth substrate. Because the optical technique requires some light to pass through the two films, their thicknesses cannot be more than a few tens of nm. It is possible to use the metal films themselves as one or both optical interferometer mirrors, but the fringes of equal chromatic order would disappear from the visible spectrum if the two

metal surfaces were brought closer together than about 1 μm (40 μin.). In that case, an alternative method of measuring separation, such as capacitance, would be required.

Environments. Experiments in the SFA can be conducted in many different liquids or vapors, as long as they are compatible with the materials of which the SFA is constructed, namely, stainless steel, silica, Kel-F, and Teflon. There is a provision to heat the chamber to around 100 °C (212 °F). Use of an appropriate heating jacket could extend the temperature range from, perhaps, -50 to 150 °C (-60 to 300 °F). At about 150 °C (300 °F), the silver layers used for interferometer mirrors degrade. This limit might be raised by using other optical coatings. The next limitation of the current design would be the maximum operating temperature of the Teflon seals, which is 250 °C (480 °F). In principle, the same or comparable techniques could be extended to operate at several hundred degrees, but, in practice, this would require a major redesign of the apparatus.

At present, the SFA is intended only to operate at or near ambient pressure. With some modifications to the seals it could be made to hold moderate vacuum, say 10^{-4} Pa (10^{-6} torr). A total redesign would be required to build a device for making comparable measurements in ultrahigh vacuum (UHV) conditions.

Preparation of Surfaces and Fluids. The first characteristic required of any solid to be investigated in the SFA is surface smoothness. This means that the solid should be smooth, compared to the range of forces of interest, if one wishes to interpret the results in a meaningful way. For van der Waals forces, this may be a few nm. For forces mediated by adsorbed polymer layers, this may be few tens of nm. For electrical double-layer forces inn water, this may be hundreds of nm. Electrostatic forces between charged surfaces in nonpolar liquids or vapors may extend for several μm.

Because the adhesion between surfaces is often dominated by very short-range forces, atomically smooth surfaces would be required to make fundamental and reproducible measurements of these. However, rough surfaces still adhere, and so measurements can be made without insisting on atomic smoothness. The drawback, in that case, is that it would be more difficult to obtain a straightforward interpretation of the results.

It is not difficult to understand that if one wants to measure surface forces between ultra-smooth, extended surfaces separated by a small number of nanometers, there are stringent requirements for cleanliness, both of the surfaces and of the environment. One speck of dust in the wrong place can spoil it all. The importance of surface cleanliness is again related to the range of force under investigation. For very short-range forces, even a monolayer of adsorbed vapor will dramatically influence results.

Considerable care must also be taken in preparing any liquid or vapor environment. Vapors must be free of dust, and liquids must be free of both particulate and molecular (for example, polymer or surfactant) contamination that can easily adsorb to solid surfaces and affect the results.

Other Measurements with the SFA. The techniques embodied in the surface force apparatus can be used to measure other properties of thin liquid or vapor films and solids at or near contact. These properties are described below.

Adsorption. The thickness of an adsorbed layer of surfactant or polymer can readily be determined by the change in surface separation before and after adsorption occurs. Compressibility of an adsorbed layer can also be measured.

Refractive Index. The optical interferometry technique can be used to measure the refractive index of very thin liquid films or adsorbed layers (Ref 10).

Viscosity. Measurement of the force when the surfaces are in relative motion allows a determination of the viscosity of ultrathin liquid films in a squeeze-film geometry (Ref 12) or simple shear (Ref 13).

Friction. A straightforward extension to lateral force measurements allows the friction between two molecularly smooth solids to be measured under controlled loads and well-determined surface separations and contact areas (Ref 14).

Atomic Force Microscopy

Following the invention of the scanning tunneling microscope (STM), which is described in a separate article by that name in this Section of the Volume, Binnig, Quate, and Gerber (Ref 15) devised an instrument which they dubbed an atomic force microscope (AFM). This instrument barrows elements from both the SFA and STM, in order to measure very small forces between a sharp tip and a flat surface (Fig. 2). Resolution in force is currently 10^{-11} N or better. Surface-

tip distances are resolved with a resolution as fine as 0.01 nm, and lateral distances along the surface are measured, typically, to 0.1 nm (see Ref 16 and the references therein). AFMs are now commercially available from a number of sources.

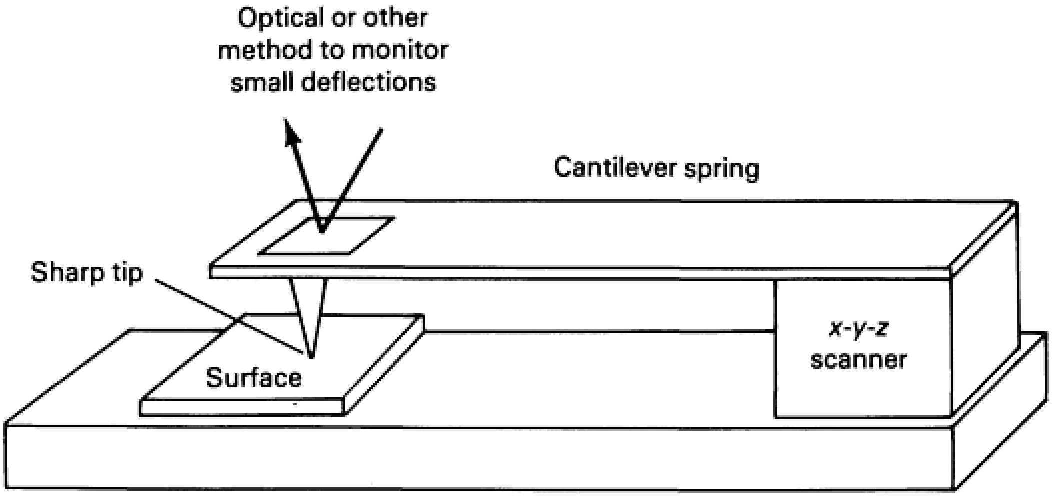


Fig. 2 Atomic force microscope, which measures the force between a sharp tip and a flat surface

Like the SFA, the AFM determines force by measuring the deflection of a cantilever spring, to which the tip is mounted. Several ingenious methods of measuring or monitoring extremely small spring deflections have been devised. The original AFM (Ref 15) used an STM tip at the back of the cantilever spring to monitor tiny movements of the spring. More recent versions have used optical interferometers or optical levers to amplify very small deflections.

The major use of the AFM to date has been as a method of obtaining atomic- or near-atomic-resolution images of surfaces that do not need to be conducting (in contrast to STM experiments). Much less work has been done on obtaining quantitative force measurements, although progress in this area is accelerating (Ref 17). In terms of measuring surface forces, the AFM has some advantages and some disadvantages, compared to the SFA technique described above.

The main advantage is the ability to measure forces between a tip and a very small region of the surface, so that surface inhomogeneities (including roughness) can be explored. A second advantage is that local surface imperfections or contamination (for example, a dust particle) can be avoided by moving to another region. In addition to forces measured out of contact, surface mechanical properties can be explored by using the AFM as a tiny indentation machine. Finally, the restrictions on materials are much less severe. Experiments could be conducted using any solid surface and any material from which a very sharp tip can be fashioned.

The major disadvantage of the AFM is that interpretation of the force is much more difficult between a tip and a surface than it is between two extended surfaces. Theories of surface forces in the tip-surface configuration are difficult at best, because the tip may be too small for continuum models to be appropriate, and too large for complete atomic calculations to be feasible. Furthermore, a proper theory would require a detailed knowledge of the tip shape and/or atomic structure, and this is rarely available. However, the field is progressing rapidly, and significant advances can be expected in the near future. There is already one impressive example of an atomic-scale computation that reproduces the features of an AFM experiment (Ref 18).

Measuring Adhesion

In light of the previous discussion, two views can be taken of adhesion measurements. One view is that to properly understand the forces and bonding that determine adhesion, an extremely careful experiment is required. This could be designated as a fundamental adhesion measurement, and the idea would be to match as closely as possible the thought experiment described in the section "Basic Concepts." For the reasons elaborated below, an exact match would be very difficult to achieve in practice.

On the other hand, adhesion is known to occur quite commonly in the real world between rough and comparatively dirty surfaces. Therefore, another view is that a practical adhesion measurement can be made if one simply wants to quantify the work of adhesion in a given engineering situation. If that is the case, then the measurement is somewhat easier than the surface force and atomic force experiments described above, because a single, comparatively large force is measured, and because surface separation does not need to be ascertained with any precision. However, even practical adhesion measurements present difficulties, because the results are very sensitive to the manner in which the two bodies are separated. These difficulties have precluded the establishment of standardized adhesion tests applicable to a wide variety of situations.

Indeed, there are so many variations on adhesion measurements (see, for example, Ref 19 and 20) that it would be impossible to describe them all fully here. Instead some general considerations are discussed, with the aim of helping readers establish sensible tests for their own particular purposes.

Fundamental Adhesion Measurements. From the earlier discussion, it is already apparent that the concept of work of adhesion includes considerable subtlety and difficulties, even in the thought experiment outlined there. Although one can blithely talk of "reversibility" in a thought experiment, that notion is particularly problematic in practice. In real adhesion measurements, processes such as surface reconstruction, interdiffusion, adsorption, chemical bonding, and either viscous or plastic flow (Ref 5) are all likely to be irreversible. This makes it very difficult to establish an unequivocal correspondence with the theoretical work of adhesion.

An ideal adhesion measurement would require a well-defined interface between two perfectly clean and atomically flat surfaces, with no adsorbate, and no interdiffusion of the two materials. The two surfaces would then need to be separated infinitely slowly in a controlled environment. These tasks would be difficult to accomplish. With these cautions in mind, the ways to conduct idealized experiments that are currently most practicable are discussed next.

Adhesion Between Curved Surfaces. To avoid problems associated with the mode of separation of two bodies (discussed further below), "fundamental" adhesion measurements can be conducted by separating two solids having curved surfaces. Examples include two spheres, a sphere and a flat, or crossed cylinders. As noted in the section "The Pull-Off Force," it is possible to make a simple association between work of adhesion (or at least, that part of it due to surface forces, W_{SF}) and the pull-off force, F_p that is the minimum tensile force required to completely separate the bodies:

$$F_p = 2\pi R W_{SF} \quad \textbf{(Eq 5)}$$

This relationship, sometimes called the DMT expression (Ref 21), is appropriate for rigid bodies that do not deform significantly upon contact. When the bodies are not completely rigid, adhesion between two curved bodies deforms them (elastically and/or plastically), and the contact point expands to a finite contact area. It was shown by Johnson, Kendall, and Roberts (JKR) (Ref 22) that if one or both of the bodies is deformable elastically, stored elastic energy is recovered in the pull-off process, and the force required for pull-off is reduced to

$$F_P = \frac{3}{2}\pi R W_{SF} \quad \textbf{(Eq 6)}$$

For example, determinations of W_{SF} have been made by measuring the pull-off force between rubber spheres (Ref 22), polyurethane, and glass (Ref 23), and crossed cylinders of mica (Ref 24), as well as measurements of the dependence of pull-off force on the relative orientation of two mica sheets (Ref 25).

If one (or both) of the solids is plastic, energy is lost during the deformation that occurs upon contact, and it is not recovered during separation. In fact, when clean metal surfaces come into contact, the surface forces alone are sufficiently strong to cause plastic deformation (Ref 26), and the simple elastic theory no longer applies. In that case, the pull-off force required to separate the bodies is greater than the JKR value (Eq 2); no upper limit can be specified.

Fracture Experiments. The well-known Griffith criterion for fracture of a monolithic brittle material equates the fracture threshold (measured as the "strain-energy release rate," or the stored elastic energy recovered as new crack is propagated through unit area) to the work of cohesion. In fact, a carefully conducted fracture experiment is as close as one can get to the ideal thought experiment described in the section "Surface Energy." If an experiment can be arranged so

that a brittle crack is propagated along the interface between two different materials, then the fracture threshold is the work of adhesion. This can be done for various material combinations, including adhesive joints and coatings (Ref 1, 19).

There is, in fact, a close analogy between brittle fracture experiments and the pull-off experiments using curved elastic bodies discussed in the preceding section. The JKR theory of pull-off (Ref 22) was derived from a fracture mechanics argument very similar to the original Griffith argument, balancing mechanical work and stored elastic energy against the interfacial energies involved in creating new surfaces. If sufficient care is taken in conducting either experiment, and attention is given to the considerations discussed below, then there is little, in principle, to choose between the two methods. The best grounds for choice would have to do with sample preparation and geometry. In particular, is it easier to prepare the requisite specimens in a curved shape or as flat, mated surfaces suitable for a fracture experiment?

History Dependence and Sample Preparation. As discussed previously, the quantity called the work of adhesion, that is, the work required to separate two bodies from contact to infinity, is a sensitive function of just where the "contact" is and the structure of the interface. Questions that need to be addressed include: Have the two solids really come into interatomic contact? Are the surfaces clean? Have they reconstructed? Have the solids annealed, cold-welded, or interdiffused? Is there still adsorbate trapped between them? What are their crystal orientations?

The answers to these questions partly depend on the properties of the two materials and their environment, and partly on the manner in which they have been prepared and brought into contact. Careful consideration must be given to these factors before meaningful and reproducible results can be obtained.

If the aim is to conduct an ideal adhesion measurement, then the two materials and any liquid or vapor environment in which the test is carried out must, at the very least, be scrupulously clean. If the aim is to design a test to measure adhesion in order to predict failure in a real engineering situation, the test conditions should be carefully chosen to match those that pertain to the real world. The adhesion measured between a given pair of metals under UHV conditions at room temperature, for example would have little to do with the adhesion that may occur between the same metals in a lubricated bearing operating at high temperature.

Rate-Dependent Effects. In many real situations, it is found that the force required to effect pull-off or crack propagation in separating two bodies depends on the rate at which the separation occurs. Rate-dependent effects can arise from viscoelasticity of one or both of the solids, viscosity of the environment, or diffusion-limited adsorption of environmental species to the newly created surfaces and possible chemical reaction with those surfaces. Generally, rate-dependent effects are always associated with dissipation of energy. Therefore, the work required to separate the bodies at anything other than infinitesimal speed is greater than the true work of adhesion.

The question of viscoelasticity in one or both of the solid bodies has been carefully examined by Greenwood and Johnson (Ref 27) and by Maugis and Barquins (Ref 23). The pull-off force becomes dependent on the rate at which it is applied, with the true pull-off force only being determined when the load is increased extremely slowly and when a very long time is allowed for the pull-off to occur. Because more rapid separation requires more force, measurements are likely to give an overestimate of the equilibrium value.

Rate effects can occurs for other reasons. For example, in some cases it takes a long time for molecules from the environment to equilibrate with the freshly exposed surfaces as a crack propagates or as the contact area between curved bodies decreases before pull-off occurs. This is well-known in the fracture mechanics field as environmentally assisted, or "slow," crack growth. Once again, unless the load is increased extremely slowly to the fracture threshold or to pull-off, the work of adhesion would be overestimated. To indicate the meaning of extremely slow, the load required to propagate a crack in glass in a humid atmosphere has been shown to be rate-dependent at crack speeds as low as 10^{-13} m/s (4×10^{-12} in./s). Ideal adhesion measurements could require some patience.

Practical Adhesion Measurements. There are many existing tests of adhesion that operate in a variety of geometries and modes of separation (Ref 1, 19, 20). A number of them are either empirical or semiempirical. Others give only a comparative test of adhesion, such as whether the adhesion between a thin film and a substrate is greater or less than the adhesion between Scotch tape and the film. There is a need for better-controlled and better-understood adhesion measurements, even at the practical level.

Practical adhesion tests involve applying a known stress to a joint or interface to determine when it fails. Failure is not always undesirable, as in the case where adhesion contributes to unwanted friction and wear. Stress can be derived from mechanical (including ultracentrifugal and ultrasonic), thermal, optical, and electromagnetic sources (Ref 20).

Apart from the difficulties outlined above, there is another complicating factor in the quest to measure the strength of an adhesive bond between two bodies. It is a common experience that the force required to separate two bodies with mating surfaces depends very much on the manner in which they are separated. For example, two microscope slides held together by a thin film of moisture are extremely difficult to separate in simple tension or even by wedging them apart, whereas they separate easily in shear. The reason for this lies partly in the fact that the work of adhesion can be done by applying a large force over a small distance (uniform tension) or a small force over a large distance (as in peeling or sliding), and partly because the fracture threshold actually depends on the mode of separation, such as tension versus shear. An adhesion test carried out in one mode might reveal little about failure in another.

Modes of Separation. Various arrangements that can be employed in adhesion tests are illustrated in Fig. 3. Detailed discussions of these can be found in texts such as Anderson *et al.* (Ref 19).

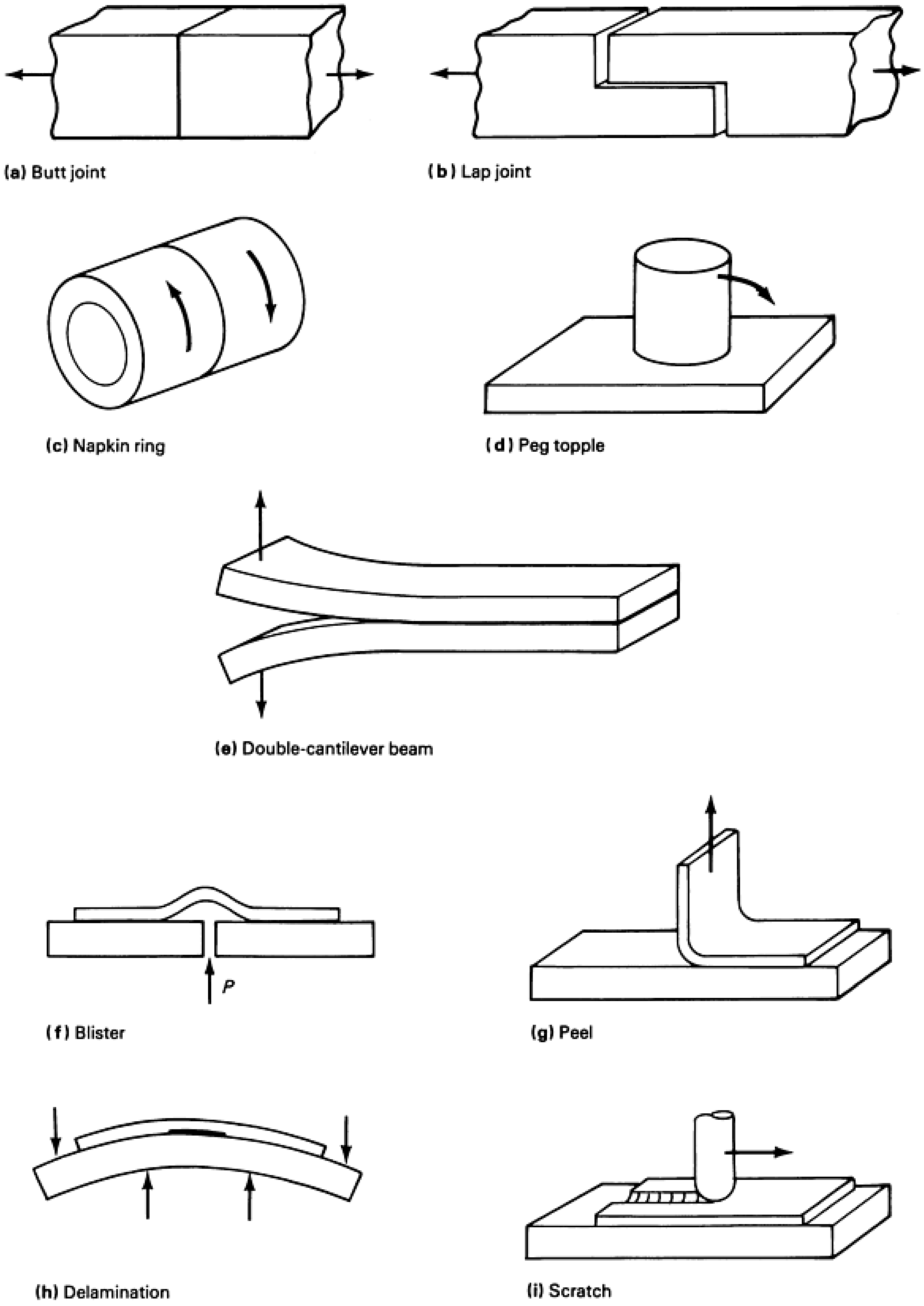


Fig. 3 Various arrangements for adhesion tests

For voluminous bodies (as opposed to films or coatings), the most obvious arrangement is the butt joint (Fig. 3a). Although it may look simple, the ease of fracture or failure can depend strongly on the presence of flaws in the joint, and on how and whether a crack/separation is initiated at the edges of the sample. A lap joint (Fig. 3b) is appropriate to test

shear strength, for example, in laminates. Care must be taken to avoid excessive bending of the beams during the test, because that introduces some tensile component. The ring shear, or "napkin ring," test (Fig. 3c) applies a more uniform shear stress to the joint, whereas the peg topple test (Fig. 3d) is closer to ideal tension, but not straightforward to analyze.

The double-cantilever beam geometry (Fig. 3e) provides an excellent fracture mechanics type of test if suitable samples can be prepared. A clever variation on this is to profile the beams in such a way that the fracture threshold is independent of the length of the crack (Ref 28), making the analysis very simple.

The tests described above are most commonly applied to testing of adhesive joints (bodies glued together with an adhesive), although they can also be used for "pure" solid-solid adhesion tests if samples can be prepared with their surfaces accurately mated together. This raises questions of surface roughness. Unless the surfaces are atomically smooth and their shapes conform precisely, the true area of contact will be less than the geometrical area of contact, and the interface will include unbonded flaws, which will serve to initiate failure. Indeed, most "flat" contacts actually involve only 8 to 50 areas of true atomic contact.

When thin films, thick films, or coatings are involved, there is another set of test geometries that can be utilized. The peel test is very common, with the force applied at various angles, not just the 90° angle shown in Fig. 3(g). Blister and delamination tests (Fig. 3f and h) can be very well controlled, and properly analyzed in terms of fracture mechanics (Ref 19, 29). The scratch test (Fig. 3i) is much more difficult to analyze, but it at least has the merit of being the only one of these configurations that tests adhesion under dynamic sliding conditions.

State of the Art

It is now possible to make accurate and well-characterized measurements of surface forces at long range, that is, surface separations from 0.5 to 100 nm, in a range of environments, and between an increasing number of materials, including mica, sapphire, glasses, metal films, and certain polymers. Information on forces between molecules and surfaces can often be gleaned by detecting adsorbed layers using the same techniques. The experiments are by no means trivial or routine, because they require a specialized, unautomated apparatus, and they are fraught with difficult problems of surface preparation and cleanliness.

Most of the measurements to date have been made at ambient temperature and pressure. They could conceivably be extended to higher temperatures and lower (or higher) pressures, but this would require some design modifications, which would add to the experimental difficulty.

Adhesion measurements present a different set of problems, because the work of adhesion is so sensitive to the precise state of the surfaces in contact and to the manner in which those surfaces are separated. The experimentalist must think carefully about exactly what he or she wishes to measure. On the one hand, fundamental measurements of adhesion under ideal conditions are inherently difficult because of the problems of obtaining perfectly clean surfaces in anything other than UHV conditions. On the other hand, more practical measurements of adhesion as it pertains to real-life applications depend very much on the mode of separation of two surfaces, making it difficult or impossible to design a single test that is appropriate for a variety of different situations.

Nevertheless, it is perfectly feasible to design adhesion tests for particular aims and applications. It simply must be accepted that the results of one test cannot easily be transferred to another situation, and this, for the time being, remains the state of the art.

References

1. K.L. Mittal, Ed., *Adhesive Joints,* Plenum Press, 1984
2. G.L. Schneberger, Ed., *Adhesives in Manufacturing,* Marcel Dekker, 1983
3. J.N. Israelachvili, *Intermolecular and Surface Forces,* Academic Press, London, 1985
4. R.G. Horn, Surfaces Forces and Their Action in Ceramic Materials, *J. Am. Ceram. Soc.,* Vol 73, 1990, p 1117-1135
5. D.H. Buckley, *Surface Effects in Adhesion, Friction, Wear and Lubrication,* Elsevier, Amsterdam, 1981
6. F.M. Orr, L.E. Scriven, and A.P. Rivas, Pendular Rings between Solids: Meniscus Properties and Capillary Force, *J. Fluid Mech.,* Vol 67, 1975, p 723-742

7. K.B. Lodge, Techniques for the Measurement of Forces Between Solids, *Adv. Colloid Interface Sci.,* Vol 19, 1983, p 27-73
8. D. Tabor and R.H.R. Winterton, The Direct Measurement of Normal and Retarded van der Waals Forces, *Proc. R. Soc. (London) A,* Vol 312, 1969, p 435-450
9. J.N. Israelachvili and G.E. Adams, Measurement of Forces between Two Mica Surfaces in Aqueous Electrolyte Solutions in the Range 0 to 100 nm, *J. Chem. Soc. Faraday Trans. I,* Vol 74, 1978, p 975-1001
10. J.N. Israelachvili, Thin Film Studies Using Multiple-Beam Interferometry, *J. Colloid Interface Sci.,* Vol 44, 1973, p 259-272
11. R.G. Horn and D.T. Smith, Analytic Solution for the Three-Layer Multiple Beam Interferometer, *Appl. Opt.,* Vol 30, 1991, p 59-65
12. D.Y.C. Chan and R.G. Horn, The Drainage of Thin Liquid Films Between Solid Surfaces, *J. Chem. Phys.,* Vol 83, 1985, p 5311-5324
13. J. Peachey, J. Van Alsten, and S. Granick, Design of an Apparatus to Measure the Shear Response of Ultrathin Liquid Films, *Rev. Sci. Instrum.,* Vol 62, 1991, p 463-473
14. A.M. Homola, J.N. Israelachvili, M.L. Gee, and P.M. McGuiggan, Measurement of and Relation between the Adhesion and Friction of Two Surfaces Separated by Molecularly Thin Liquid Films, *J. Tribol.,* Vol 111, 1989, p 675-682
15. G. Binnig, C.F. Quate, and C. Gerber, Atomic Force Microscope, *Phys. Rev. Lett.,* Vol 56, 1986, p 930-933
16. D. Rugar and P. Hansma, Atomic Force Microscopy, *Phys. Today,* Oct 1990, p 23-30
17. N.A. Burnham and R.J. Colton, Force Microscopy, *Scanning Tunneling Microscopy: Theory and Application,* D.A. Bonnell, Ed., VCH Publishers, 1991
18. U. Landman, W.D. Luedtke, N.A. Burnham, and R.J. Colton, Atomistic Mechanisms and Dynamics of Adhesion, Nanoindentation, and Fracture, *Science,* Vol 248, 1990, p 454-461
19. G.P. Anderson, S.J. Bennett, and K.L. DeVries, *Analysis and Testing of Adhesive Bonds,* Academic Press, 1977
20. K.L. Mittal, Ed., *Adhesion Measurement of Thin Films, Thick Films and Bulk Coatings,* STP 640, ASTM, 1978
21. B.V. Derjaguin, V.M. Muller, and Yu.P. Toporov, Effect of Contact Deformations on the Adhesion of Particles, *J. Colloid Interface Sci.,* Vol 53, 1975, p 314-326
22. K.L. Johnson, K. Kendall, and A.D. Roberts, Surface Energy and the Contact of Elastic Solids, *Proc. R. Soc. (London) A,* Vol 324, 1971, p 301-313
23. D. Maugis and M. Barquins, Fracture Mechanics and the Adherence of Viscoelastic Bodies, *J. Phys. D, Appl. Phys.,* Vol 11, 1978, p 1989-2023
24. R.G. Horn, J.N. Israelachvili, and F. Pribac, Measurement of the Deformation and Adhesion of Solids in Contact, *J. Colloid Interface Sci.,* Vol 115, 1987, p 480-492
25. P.M. McGuiggan and J.N. Israelachvili, Adhesion and Short-Range Forces between Surfaces: Effects of Surface Lattice Mismatch, *J. Mater. Res.,* Vol 5, 1990, p 2232-2243
26. J.B. Pethica and D. Tabor, Contact of Characterised Metal Surfaces at Very Low Loads: Deformation and Adhesion, *Surf. Sci.,* Vol 89, 1979, p 182-190
27. J.A. Greenwood and K.L. Johnson, The Mechanics of Adhesion of Viscoelastic Solids, *Philos. Mag. A,* Vol 43, 1981, p 697-711
28. W.D. Bascom, P.F. Becher, J.L. Bitner, and J.S. Murday, Use of Fracture Mechanics Concepts in Testing of Film Adhesion, *Adhesion Measurement of Thin Films, Thick Films and Bulk Coatings,* STP 640, K.L. Mittal, Ed., ASTM, 1978, p 63-81
29. D.B. Marshall and A.G. Evans, Measurement of Adherence of Residually Stressed Thin Films by Indentation. Mechanics of Interface Delamination, *J. Appl. Phys.,* Vol 56, 1984, p 2632-2638

Characterization of Surfaces by Acoustic Imaging Techniques

W. Arnold, U. Netzelmann, and S. Pangraz, Fraunhofer Institute for Nondestructive Testing (Federal Republic of Germany)

Introduction

SCANNING ACOUSTIC MICROSCOPY (SAM), high-frequency acoustic imaging (HAIM), and scanning laser acoustic microscopy (SLAM) are based on the interaction of high-frequency acoustic waves with the sample material. The contrast of acoustic images is primarily determined by the elastic properties of the object under investigation. Because acoustic waves can also penetrate opaque materials, it is possible to image internal structures of the sample, to look for defects, to evaluate homogeneity, and to gain information on elastic parameters. These properties make acoustic microscopy and high-frequency acoustics important tools for nondestructive evaluation (NDE) and materials characterization. Acoustic imaging at low frequencies ($\leq$10 MHz) is generally used for defect detection--an application of great practical importance. Whereas theories underlying acoustical imaging comprise both regimes, imaging at higher frequencies reveals more details about the microstructure, because scattering, absorption, and dispersion caused by defect agglomerates, second phases, dislocations, inhomogeneities, and plastic deformation become more significant. This can be exploited in materials science and technology.

Various types of commercial and experimental equipment on the market are suitable for this task. The SAM microscope is the analog of the optical microscope, because it uses focused waves (Ref 1, 2) covering the frequency range from 100 MHz to 2 GHz. HAIM equipment operates in the frequency range from 10 to 100 MHz and is generally employed to characterize advanced materials where small defects are to be detected (Ref 3). In SLAM, plane ultrasonic waves traveling through the sample (with a frequency ranging from approximately 30 to 500 MHz) are used for imaging (Ref 4). By scattering and absorption, the internal structures of the sample modify the sound field according to their elastic and geometric properties and thus can be made visible. This article presents typical examples of applications for the three types of instruments, outlining the physics of the contrast mechanism and the applications for these instruments in materials science, NDE, and tribology.

Acknowledgement

This work was supported by grants from the European Community and by the German Ministry of Research and Technology within the Material Research Programme.

Scanning Acoustic Microscopy

Principles of SAM

In SAM, a focused ultrasonic beam is used for imaging. Two types of SAMs are commonly used: transmission mode and reflection mode. In commercial models, the reflection mode is dominant. Its working principle is shown in Fig. 1. The ultrasonic beam is excited by a transducer such as a ZnO film sputtered on the plane surface of a sapphire delay rod. On the opposite side of the delay rod, a spherical curvature is carved into it with high precision. The ultrasonic beam is then refracted by the curvature and is focused onto the surface of the sample to be examined, provided a coupling medium is used whose sound velocity (v_{Lc}) is smaller than that of sapphire (v_{Ls}), a material often used as a delay rod (v_{Ls} = 11.2 mm/μs). Distilled water is usually used as coupling medium (v_{Lc} = 1.5 mm/μs). The sapphire delay rod is cut into the *c*-direction because the ultrasonic absorption is extremely small in this material, and in particular in this direction. It is possible to propagate ultrasonic waves in sapphire of $\leq$4 GHz frequency at room temperature with an acceptable signal/noise ratio.

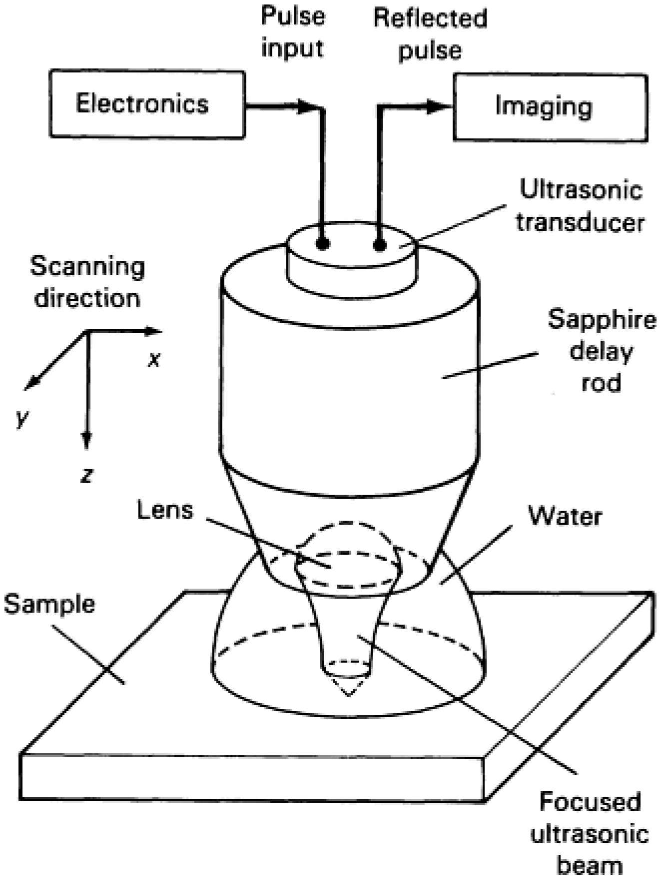


Fig. 1 Schematic showing key components of the SAM. The spherical interface between the sapphire rod and the coupling medium (usually water) acts as a lens with a large refractive index (~7.5) for ultrasonic waves traveling down the delay rod. The focused ultrasonic beam is reflected back into the lens, and an image is built up by scanning the lens in the *x*- and *y*-direction and displaying a signal on a screen proportional to the amplitude back-reflected.

The refractive index of such a one-lens system is n = 11.2/1.5 ≈7.5, much larger than that of objectives in optical microscopes. The large refractive index of the lens allows the SAM to be operated with only a one-lens objective, because spherical aberration is proportional to n^{-2} and hence is negligible in SAM. Chromatic aberration is not important, because the transducer is excited by a carrier signal with at least 10 cycles.

The focal length of an acoustic lens made of sapphire is f = 1.15r, where r is the radius of curvature of the lens. The resolution of such an instrument is approximately λ, the wavelength of sound in the coupling medium. For a frequency of 2 GHz, λ= 0.75 μm in water. Technical details such as electrical and mechanical impedance matching are important in the design and manufacture of a lens (Ref 1). The acoustic lens is scanned over the surface of the sample; acoustic energy is reflected back into the lens, is again detected by the transducer, and is converted into a radio frequency (rf) signal that is amplified, displayed on a screen, and stored in a computer after digitization. Because the signal carrying the information from the surface under inspection is time-delayed, it is gated out by a compatible electronic circuit. The reflection coefficient depends on the local elastic modulus, density, and, to a certain extent, ultrasonic absorption. Figures 2(a) and 2(b) show the images of a Vickers indentation in a metal and in a ceramic material, respectively. In the metal, the contrast change around the indentation caused by plastic deformation is visible. In the ceramic material, the cracks emanating from the corners of the indentation are clearly visible. This is because cracks disrupt the flow of acoustic energy because of the atomic disbond; that is, the crack opening displacement can be much smaller than the wavelength and the crack will still be visible.

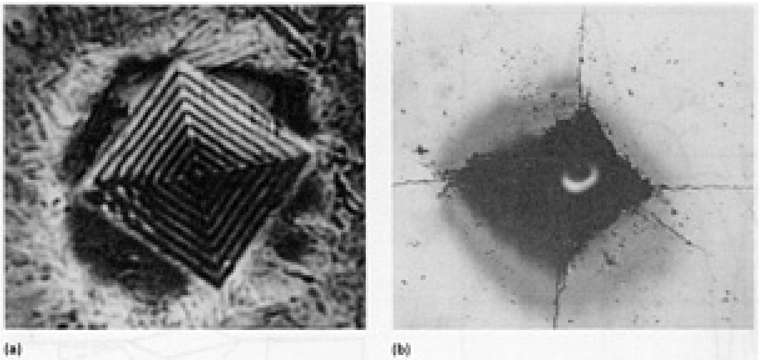


Fig. 2 (a) SAM images of Vickers indentations in two selected materials. Indentation in a steel taken at a frequency of 1.3 GHz. The change in signal surrounding the image of the indentation is caused by plastic deformation. The lines are caused by the change in contrast due to the slope of the pyramidal indentation. 913×. Courtesy of S. Ganz, Leica. (b) Indentation in SiC ceramic taken at a frequency of 1 GHz. Cracks are visible emanating from the corners of the indentation. Bulging of the material (and the possible presence of subsurface lateral cracks) changes the contrast close to the image of the indentation. 160×

There is a unique feature central to the understanding of the contrast mechanism of the SAM. The opening angle of the acoustic lenses includes the critical angle for surface-wave generation (Fig. 3). Because of the boundary condition (coupling medium instead of vacuum), the surface wave is not pure mode but rather a leaky surface wave that radiates back into the lens and interferes with all other rays. The resulting interference pattern is called a $V(z)$-curve because it becomes measurable only after defocusing the lens toward the surface by a distance z (Ref 5). The distance Δz between the maxima (or the minima) of the $V(z)$-curve is given by (Ref 2):

$$\Delta z = v_{Lc}/[2f(1 - \cos \theta_R)] \qquad \textbf{(Eq 1)}$$

with $\sin \theta_R = v_{Lc}/v_R$. Here, v_R is the Rayleigh velocity of the material under inspection, f is the operating frequency of the SAM, and θ_R is the critical angle.

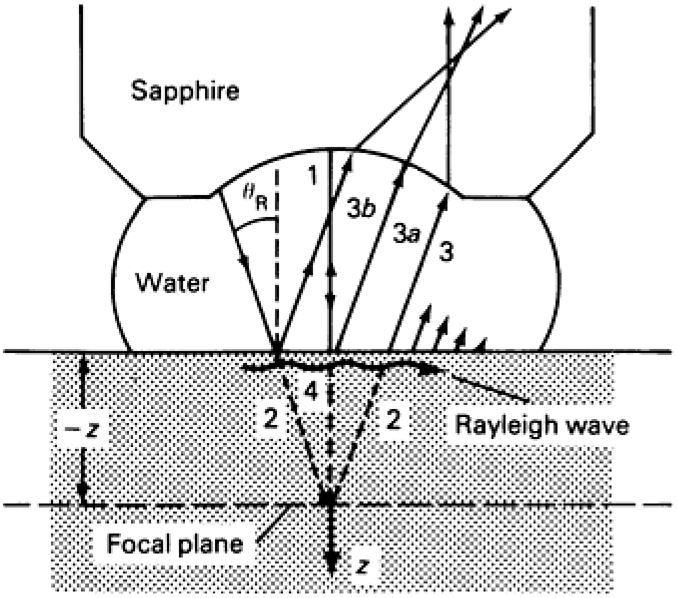


Fig. 3 Principle of Rayleigh wave generation by SAM in the framework of geometric acoustics. The ray incident under the critical angle, θ_R, generates surface waves (Rayleigh waves), because their wave fronts match one another. The central beam (ray 1) interferes with ray 3 because of their phase delay, which depends on the

actual defocus distance, *z,* thus causing the *V*(*z*)-curve. Ray 3 is caused by reradiation of the surface wave back into the water. Rays 3*a* and 3*b* do not contribute, because they are refracted such that they cannot reach the transducer. The "focal plane" is only meant for clarification of the defocus distance, *z.* In reality, the rays are refracted at the surface of the specimen under investigation.

The amount of defocusing set for the lens in a given SAM image determines the relative contrast of the features observed. Further defocusing may result in a total contrast reversal (Ref 2), considering that each given point in a SAM image exhibits a different $V(z)$-curve depending on the local elastic properties. In turn, measuring $V(z)$-curves allows determination of surface elastic modulus with a spatial resolution ranging from 5 to ~50 μm (given by the spot size of the acoustic beam at the surface), depending on the amount of defocusing and the material examined. Therefore, $V(z)$-curves are also called acoustic material signatures (AMS) (Ref 5) (see Fig. 4). In anisotropic materials, analysis of the $V(z)$-curve becomes more complicated (Ref 6). By using cylindrical lenses, it is possible to measure the Rayleigh velocity as a function of angular direction at the surface of a material (Ref 7). Finally, the scattering of the surface waves by inhomogeneities such as cracks, porosity, and inclusions is also an important contribution to the contrast in the images obtained, as can be see in Fig. 2.

Applications of SAM

Numerous applications of SAM have been reported in the technical literature that also illuminate the contrast mechanism. This discussion will be restricted to a few relevant examples regarding the use of SAM to measure the surface wave velocity, v_R.

Thin-Film Thickness Measurements. The excitation of surface waves by the acoustic lens may be exploited to measure the thickness of thin films coated on a base material (Ref 8) and to characterize their bond strength (Ref 9). The presence of a film whose elastic properties are different from those of the base material produces a composite in which the surface acoustic wave (SAW) travels and then becomes dispersive. It is possible to derive the film thickness from the measured dispersion curve, provided either a calibration curve is used or the propagation of the surface wave is modeled using the elastic data of the film and the base material (Ref 10). Continuous waves (cw) are also used to determine the velocity of SAWs via $V(z)$-curves in layered media (Ref 11). Here, a continuous change of the operating frequency allows the dispersion diagrams of several material combinations to be determined. The $V(z)$-curves obtained by cw excitation are similar to those of burst-generated $V(z)$-curves (Ref 11).

Evaluation of Surface Conditions Caused by Machining. Material topography is an important factor in the contrast of SAM images. This is caused by the phase difference of the various angular rays generated by the acoustic lens after reflection and by the phase-sensitive nature of the transducer. In fact, topography can be drastically enhanced by phase-sensitive electronic detection systems in SAM, allowing measurement of very tiny steps on a surface (less than 10 nm) (Ref 12) and the local stress distribution by exploiting the acoustoelastic effect (Ref 5, 13, 14), which describes the stress dependence of sound velocity in terms of higher-order elastic constants. Likewise, phase-sensitive detection enhances the capability of SAM for detecting small defects and cracks in machined surfaces; this process has been used for quality assurance of ceramic ball bearings (Ref 15).

Machining damage in metals alters the elastic properties of a surface layer to a depth of a few tens of micrometers. This may be caused by twinning of the crystallographic structure and by the introduction of dislocations. In turn, this lowers the surface wave velocity, which can be measured with high precision using the $V(z)$ technique. Such a measurement was undertaken in machined beryllium (Ref 5), and a 3% decrease in v_R relative to the undamaged material was observed (Fig. 4), similar to the decrease observed in wear studies of alumina surfaces (Ref 14).

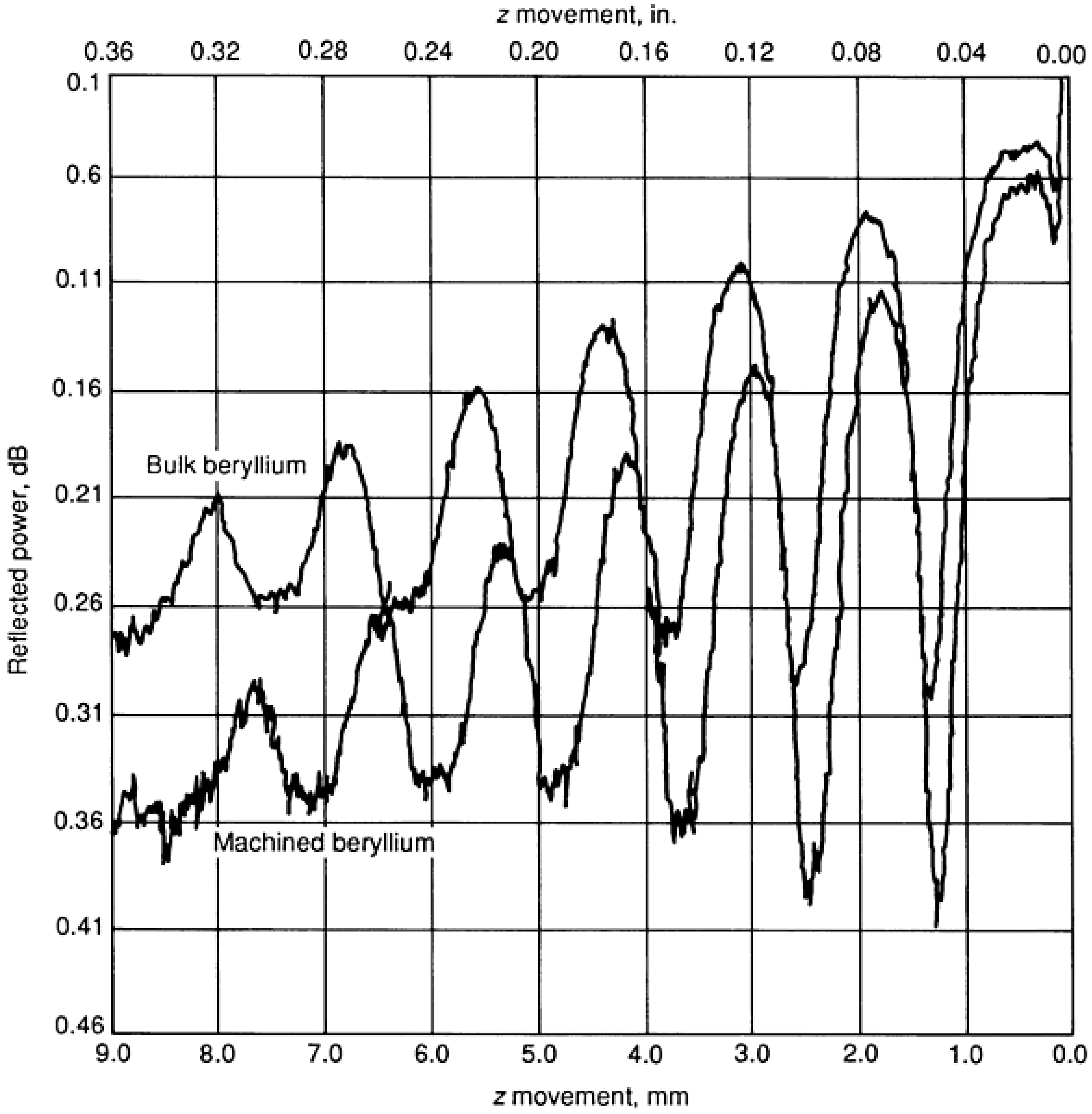


Fig. 4 *V*(*z*)-curves of machined beryllium surfaces taken at 35 MHz. The Rayleigh velocity, v_R, of the unmachined beryllium surface is 8.07 mm/μs; the v_R of the machined beryllium surface is 7.82 mm/μs. The number of maxima obtainable in a *V*(*z*)-curve depends on the ratio of the focal distance of the lens to the wavelength, λ_R, of the surface wave. Source: Ref 5

Extensive studies of damage and wear on ceramic surfaces have been carried out using SAM (Ref 16). First, the change of contrast observed in SAM images caused by the presence of subsurface lateral cracks in hot-pressed silicon nitride and steel ball bearings was examined (Ref 17) (Fig. 5(a) and 5(b)). Second, the fracture toughness, K_c, of the material was determined by measuring the lateral extent, C_1, of the contrast change close to the indentation using the expression:

$$C_1 = a\left[\frac{(E/H)^{3/4}}{(K_c H^{1/4})}\right]^{1/2} \cdot P^{5/8} \quad \text{(Eq 2)}$$

where a is a constant determined by the shape of the indenter, *E* is the Young's modulus, *H* is the hardness, and *P* is the indentation load. Furthermore, the sliding wear resistance of thin TiN films coated on a die steel was monitored by exploiting the scattering of the Rayleigh waves at surface and subsurface cracks. The onset of wearing of the TiN film as a function of the number of sliding cycles was clearly established in the SAM images (Fig. 6).

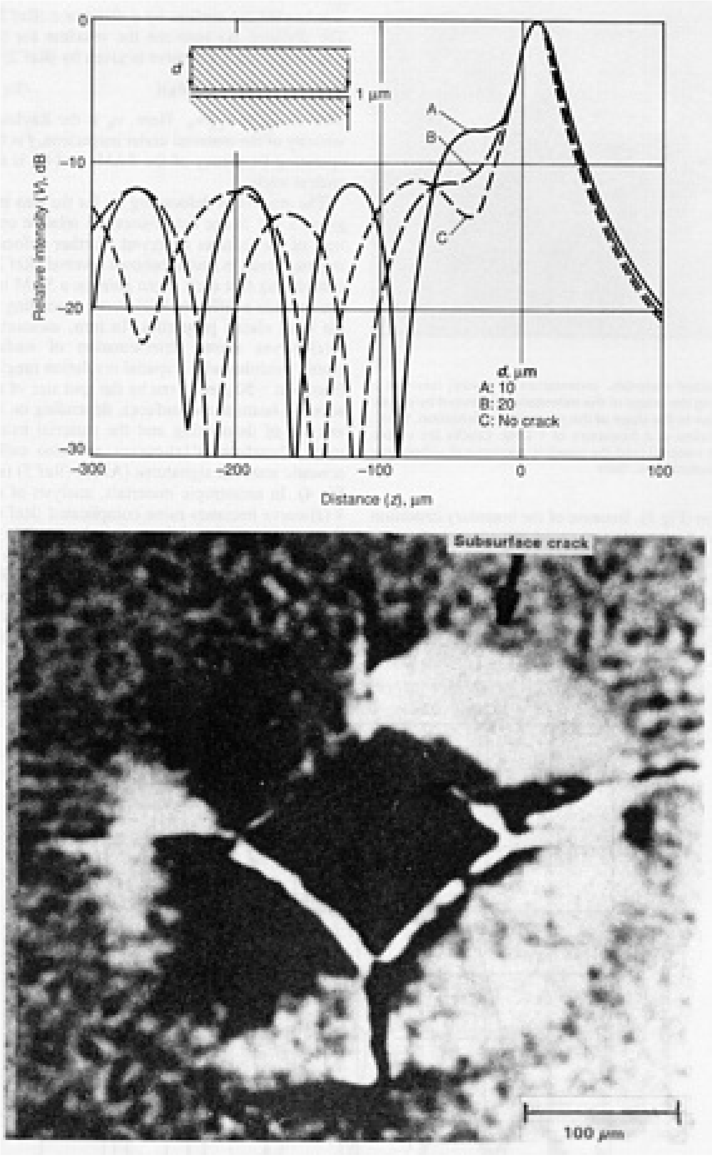


Fig. 5(a) SAM image contrast generated by surface defects in HPSN and steel materials. Top: Effect of a subsurface lateral crack present in the depth, *d,* in hot-pressed SiN on the *V*(*z*)-curve. Bottom: Change in *V*(*z*)-curve attributed to lateral cracks produces the change in contrast obtained close to the indentation. Measuring frequency was 200 MHz; half-opening angle of the lens was 60 °. See Fig. 5(b). Source: Ref 16

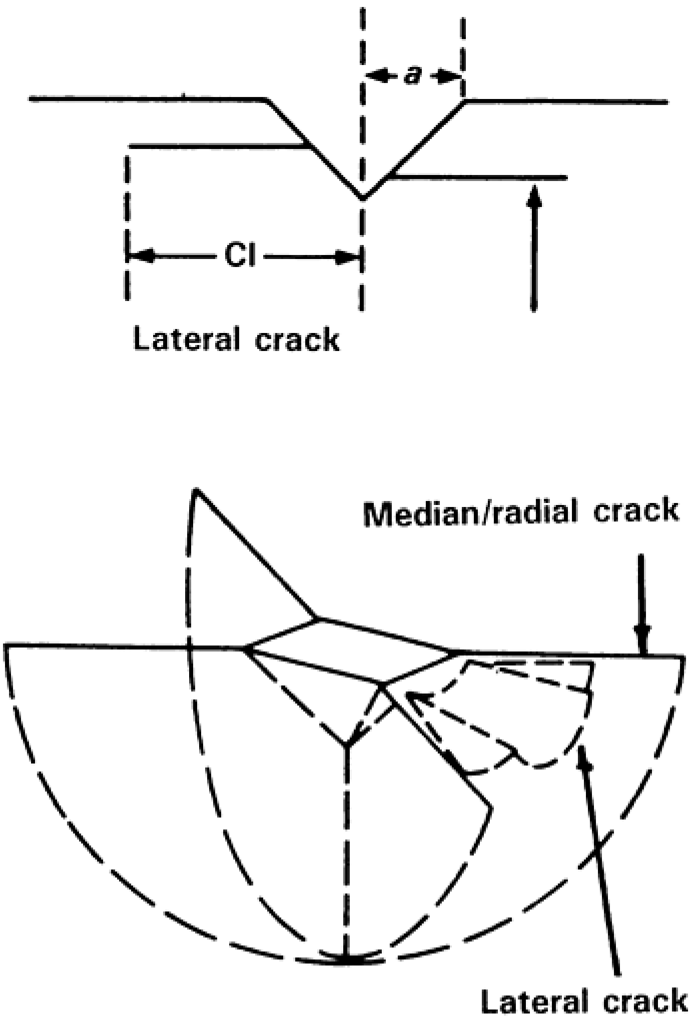


Fig. 5(b) Schematic showing detail of SAM image contrast generated by surface defects in HPSN and steel materials. Top: Lateral crack in HPSN. Bottom: Median/radial crack and a lateral crack in a steel ball bearing. See Fig. 5(a). Source: Ref 16

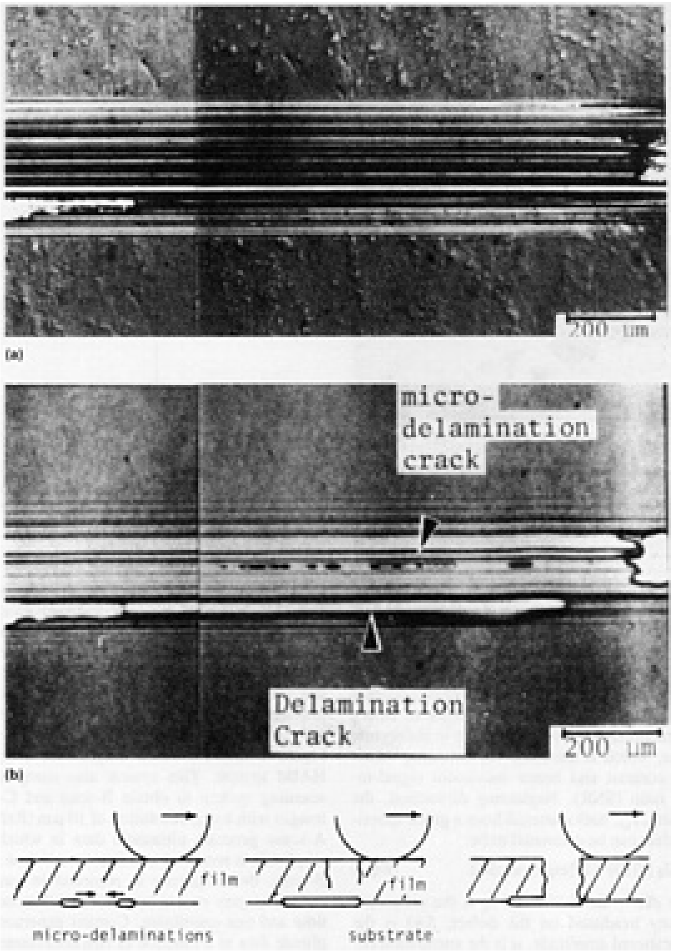


Fig. 6 Sliding damage incurred by a TiN film on a steel sample. (a) Optical image. (b) SAM image. A steel ball of 2.4 mm (0.094 in.) was slid over the film with a load of 4.9 N (1.1 lbf) (typically, 1000 repetitions are completed) and the onset of wear was observed. Microdelaminations develop that eventually grow into larger delaminations. These microdelaminations become visible in SAM images. Source: Ref 16

Machining damage in ceramics can also be evaluated by a different procedure using SAM. Common techniques for machining structural ceramics use alternating grinding (AG) and creep-feed grinding (CFG) to remove surface material. Parameters that can be varied in these processes are in-feed, feed-speed, and cutting speed, resulting in a given material removal rate. Zirconia (ZrO_2) and silicon-infiltrated silicon carbide (Si/SiC) were examined by taking SAM images after machining. A typical image is shown in Fig. 7(a). The number of cracklike indications, *N*, perpendicular to the grinding grooves per area were counted as a function of machining parameters (Table 1), exploiting the sensitivity of SAM to cracks present in a surface layer. *N* was clearly minimal in case of CFG (Fig. 7b).

Table 1 Recommended machining parameters to ensure the production of ceramic components with the least possible surface damage

Machining process	Sample groove location	In-feed		Feed speed		Cutting speed		Material removal rate (MRR) per unit wheel width	
		mm	**in.**	**mm/s**	**in./s**	**m/s**	**ft/s**	**mm^2/s**	**$(in.^2/s) \times 10^{-3}$**
Alternating grinding	N_1	0.03	0.0012	167	6.6	25	82	5	8
	N_2	0.03	0.0012	84	3.3	25	82	2.5	3.9

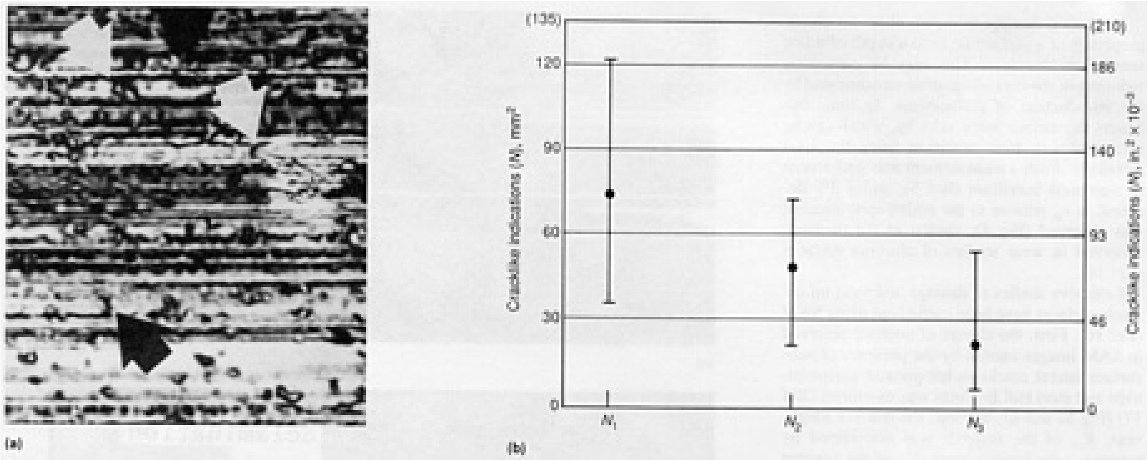


Fig. 7 Surface damage (in terms of cracklike indications) in ceramics relative to machining parameters and processes. (a) SAM image of an AG ZrO_2 sample with cracklike indications perpendicular to the grinding grooves was taken at a frequency of 200 MHz. The image size area is 780 × 740 μm^2. (b) Numbers of detected cracklike indications of ZrO_2 samples machined by AG (N_1 and N_2) and CFG (N_3) methods. The bars indicate over which margin the number of cracklike indications vary from sample to sample. The average value is marked by a dot.

Such cracks may cause failure after mechanical loading of the component. This was corroborated by measuring the bending strength of samples made of the same material and subjected to the same machining procedure. The bending strength was higher in the case of CFG (Ref 18). This evaluation technique is similar to one used to evaluate wear in a sliding experiment on hot-pressed SiN and partially stabilized ZrO_2 (Ref 16).

Furthermore, SAM has been used to evaluate corrosion attack on lacquered sheet. Again, this is based on the sensitivity of the contrast in SAM images on the elastic boundary conditions between the adhering film and the base material (Ref 14). Finally, plastic deformation occurring at grain boundaries has been observed by SAM after severely stressing polycrystalline steel containing 3% Si (Ref 19).

High-Frequency Acoustic Imaging

Principles of HAIM

The efficient detection of defects by ultrasound requires that the wavelength be comparable to the defect size. For example, because of fracture mechanics considerations, it is necessary to detect defects smaller than 100 μm in engineering ceramics and other high-strength materials. Because of the large sound velocities in these materials, use of high-frequency acoustics is required. However, the ultrasonic attenuation in the material examined may become quite large at frequencies beyond 100 MHz, limiting the penetration depth. Furthermore, the probes employed have an opening angle smaller than the critical angle θ_R, avoiding the generation of surface waves typical for SAM and hence yielding deeper penetration into the material. Otherwise, the mechanisms contributing to the contrast in SAM and HAIM are very similar.

Depending on the ratio of the defect size, *a,* to the employed ultrasonic wavelength, λ, three regimes can be distinguished that are important for defect detection: Rayleigh scattering ($a/\lambda \ll 1$), resonance scattering ($a/\lambda \sim 1$), and geometric scattering ($a/\lambda \gg 1$). As mentioned previously, the attenuation limits the propagation of ultrasonic waves in engineering materials at frequencies much above 50 MHz. Use of 50 MHz waves in ceramics and other materials with sound velocities above 6 mm/μs keeps detection in the Rayleigh regime for the order of defect sizes mentioned above, rather than in the regime $\lambda \sim a$, which is desirable for obtaining maximum contrast and hence maximum signal-to-noise ratio (SNR). Neglecting diffraction, the intensity, I_{sc}, backscattered from a given spherical defect can be estimated to be:

$$I_{sc} = I_0 \, [f(\pi)]^2 \pi a^2 [\exp(-\alpha r)] d\Omega \qquad \textbf{(Eq 3)}$$

where $d\Omega = \Delta F/4\pi r^2$, and I_0 is the ultrasonic intensity irradiated on the defect, $f(\pi)$ is the backscattered amplitude, α is the attenuation coefficient in the material, ΔF is the area of the transducer, r is the depth of the defect, and $d\Omega$ is the solid angle subtended by the probe as seen from the defect. Because $f(\pi) \propto a^2 f^2$ in the Rayleigh regime, $I_{sc}/I_0 \propto a^6 f^4$ (f is the ultrasonic frequency)! This means that very efficient transducers with a center frequency as high as possible must be employed in order to obtain a sufficiently high SNR for a given excitation voltage of the transducer. For other defect shapes, expressions similar to Eq 2 also hold true (Ref 20).

Therefore, the electronic systems and probes in a HAIM system must be designed such that the highest SNR is obtained and losses are absolutely minimized. In the setup currently used by the authors, the detection limit for defects is 30 μm for inclusions of a few millimeters depth, provided the ultrasonic attenuation in the material examined is less than 1 dB/cm at 50 MHz. Figure 8 shows a block diagram of a typical HAIM system. This system also comprises a scanning system to obtain B-scan and C-scan images with a step-resolution of 10 μm (Ref 21). A-scans generate ultrasonic data in which the amplitude is recorded as a function of time. In a B-scan, the amplitude is recorded in varying shades of gray or a color scale as a function of time and one coordinate. C-scans generate amplitude data as a function of two coordinates. In general, the maximum amplitude in C-scans is recorded for the image built up within a preset gate having a time delay that defines the time-of-flight of the signal and, hence, the depth of its origin within the sample. After rectification, the portion of interest of an A-scan is cut out by a gate and the signal strength within this gate is used to build up the image. The authors' system has been used to detect and evaluate defects, lack of adhesion between two different materials, homogeneity, and surface damage of components. Details about the design of the focusing probes can be found elsewhere (Ref 22). Various electronic systems are used for HAIM.

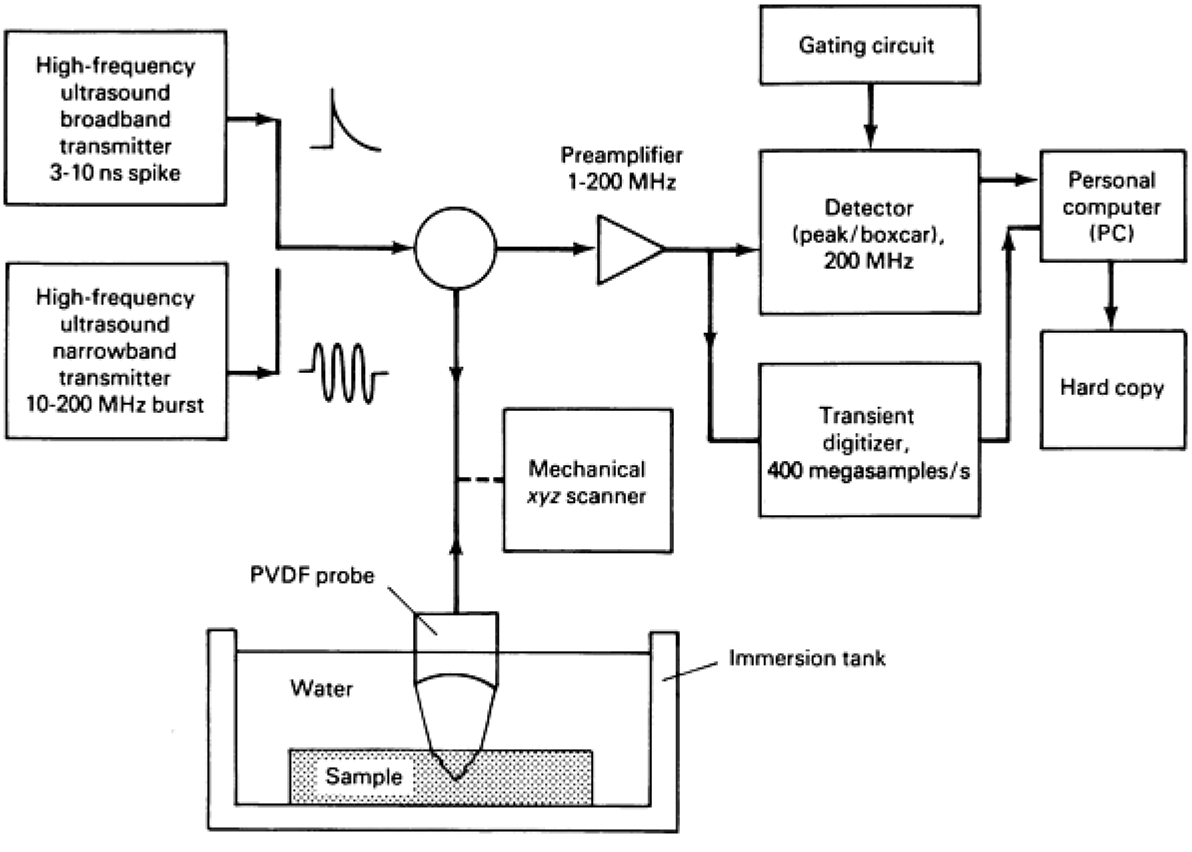


Fig. 8 Block diagram of a typical HAIM system. A transmitter excites the transducer, in this case a polyvinylidene-difluoride (PVDF) transducer. In order to obtain sufficient spatial resolution, pulses of less than 100 ns are needed. A large bandwidth for both the electronics and the transducer are then necessary. The transducer employed may be excited by either an exponentially decaying step-pulse (broadband excitation), typical in NDE electronics, or by an rf carrier pulse (narrowband excitation). The component is scanned by either an *xyz* scanning system or by a robotic system. Source: Ref 21

Applications of HAIM

High-frequency acoustic imaging is used to test bonding interfaces (Fig. 9). The typical sample used for adhesion tests for biomedical applications consists of a metal slab (15 × 10 × 2 mm, or 0.59 × 0.4 × 0.08 in.) bonded by a glue to a plastic slab of the same size. Using a 50 MHz focusing probe with broadband excitation by a spike pulse, ultrasound was sent through the surface of the plastic slab and the backscattered echo from the bonding interface was detected by a peak detector. The C-scan (16 × 16 mm, or $\frac{5}{8} \times \frac{5}{8}$ in.) shows typical distribution patterns of adhesive in the interface. Areas of large change of the acoustic impedance appear as bright colors (light in gray scale), low reflecting areas appear as dark colors (dark in gray scale). An enhanced sensitivity to surface damage can be obtained by radiating the acoustic energy under an oblique angle to generate surface waves. Surface wave scattering by defects is, therefore, the dominant source of contrast in a HAIM image, just as it is in SAM images.

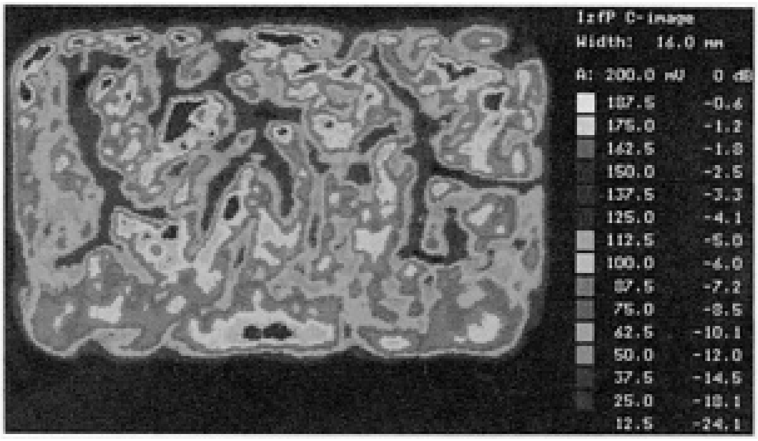


Fig. 9 C-scan image of a bonded structure obtained by high-frequency acoustic imaging. Center frequency of the probe was 50 MHz. The color scale (gray scale) is calibrated in relative intensity (dB). The width of the image is 16 mm. Original image is in color. Further details are explained in text.

Scanning Laser Acoustic Microscopy

Principles of SLAM

The operating principle of a SLAM is outlined in Fig. 10. A sample is insonified under a certain angle with respect to the surface of the sample. In a homogeneous sample, the ultrasound causes a ripple of the surface, off which a laser beam is reflected. The spatial and temporal periodic displacements of the surface cause the laser beam to be partially diffracted and frequency to be shifted by the Doppler effect. By a knife edge, one diffraction order is blocked. This then leads to an alternating current (ac) in the photodiode, because its output contains a mixing product between the undiffracted zero order and the still-present diffracted part. The frequency of the ac component is equal to the sound frequency, and its magnitude is proportional to the sound amplitude.

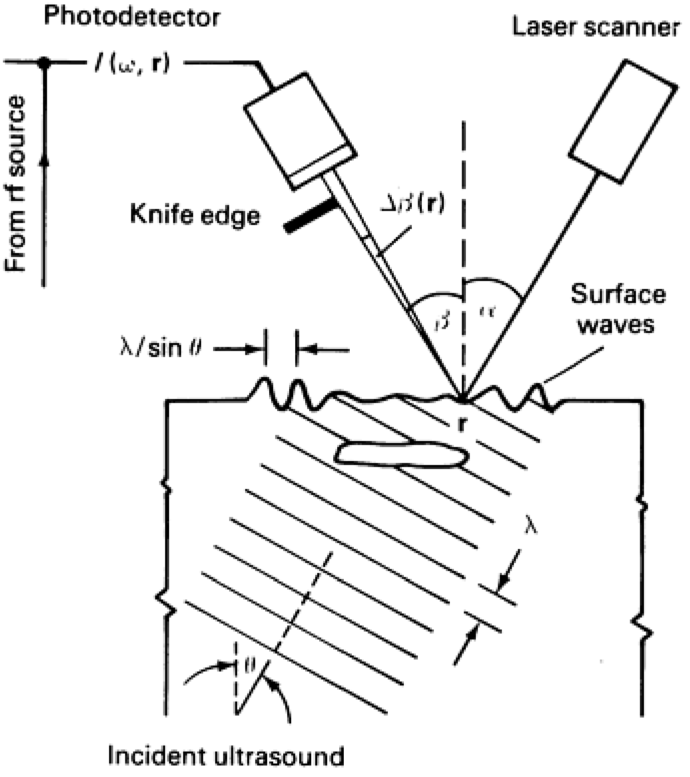


Fig. 10 Schematic showing key components and parameters of SLAM. The sound waves traveling through the sample under an angle, θ, generate a surface ripple with a wavelength of λ/sin θ. The soundwaves are detected by an optical knife-edge device in which the angle of reflection, β, is modulated by an amount, $\Delta\beta$, depending on the local amplitude of the surface ripple, according to the amount of scattering by defects present in the sample. The images obtained are acoustic holograms equivalent to Gabor holography in its original form. A quadrature receiver is used to detect the phase of the image required for holographic reconstruction. Here, the reference phase necessary for quadratic detection is obtained from the rf source driving the transducer. The images are obtained in real-time (that is, with 25 frames/second).

This technique makes possible the detection of coherent surface waves of extremely small amplitude ($\sim 10^{-6}$ nm/$\sqrt{Hz}$ bandwidth). When pores, inclusions, and cracks are present in the sample, the sound wave is scattered by these defects, which in turn become visible as a modulation of the otherwise homogeneous surface ripple. By rastering the laser beam over the surface of the sample, this modulation can be measured and displayed on a television screen (the rate of image buildup is the TV rate). The resolution in such images is given by the wavelength of the ultrasound and is typically 50 μm in most solid materials at 100 MHz.

It is obvious that the images obtained by SLAM are acoustic shadowgraphs, provided the size of the imaged structure is large compared with the wavelength. If the size becomes comparable to λ, diffraction patterns are obtained. If the surface of the sample is not optically reflective, the dynamic ripple caused by the sound field is then transmitted to a light-reflecting layer that is coupled acoustically to the sample surface by water.

Reconstruction of Images by Holography

In addition to the simple detection of defects in a given sample, SLAM techniques can be used to study their characterization and sizing. In general, this is a complex problem, because the SLAM provides two-dimensional images of a three-dimensional defect geometry deblurred by diffraction effects. Therefore, a defect appear much larger than it really is. It is possible, however, to detect both the amplitude and phases of the acoustic field in a SLAM image, allowing reconstruction of the defect by acoustic holography techniques (Ref 23).

Acoustic holography is a two-step process. First, the amplitude and phase of the acoustic field emanating from an insonified object are detected in a plane adjacent to its surface. Second, from these field data, the scattered field in the

defect plane, Φ, can then be reconstructed with a resolution of 1 wavelength. The relation between the fields $\Phi(x,y,0)$ and $\Phi(x,y,\Delta z)$ in two parallel planes in the sample under investigation, separated by a distance Δz, is given by a linear filtering process (Ref 24):

$$\tilde{\Phi}(k_x,k_y,k\Delta z) = \tilde{\Phi}_0\,(k_x,k_y,0)\cdot \exp\,(ik_z\cdot \Delta z) \qquad \textbf{(Eq 4)}$$

where $k_z = \sqrt{k^2-k_x^2-k_y^2}$ and $\tilde{\Phi}$ is the corresponding fields in k-space. k_x and k_y can be interpreted as the x- and y-component of the wave vector of a plane wave with amplitude Φ_0. Thus, the field distributions between the detection plane and the object can be obtained by calculating the two-dimensional Fourier transform of the field Φ, multiplying it with the filter function exp(- $k_z\cdot\Delta z$), and Fourier back-transforming it into spatial coordinates. This process is called back-propagation. Because outside $k_x^2+k_y^2=k^2$ the filter function is steeply increasing, the filtering is restricted to the innerface of the circle with radius k so that noise is reduced. The SNR in the back-propagated image can be further enhanced by deconvoluting the field data with the transfer function of the laser detection scheme employed in the SLAM, resulting in a total improvement of the SNR of approximately 10 dB compared with the original image.

Figure 11 shows the image of an iron inclusion in an Si/SiC bending bar obtained at an ultrasonic frequency of 100 MHz (Ref 24). The diffraction of the sound-field waves at the inclusion causes concentric ring patterns. Such images are typical for SLAM. However, by subsequently calculating the field distributions in planes of increasing depth, an image with an apparently optimal defect contrast is obtained that corresponds to the depth of the defect.

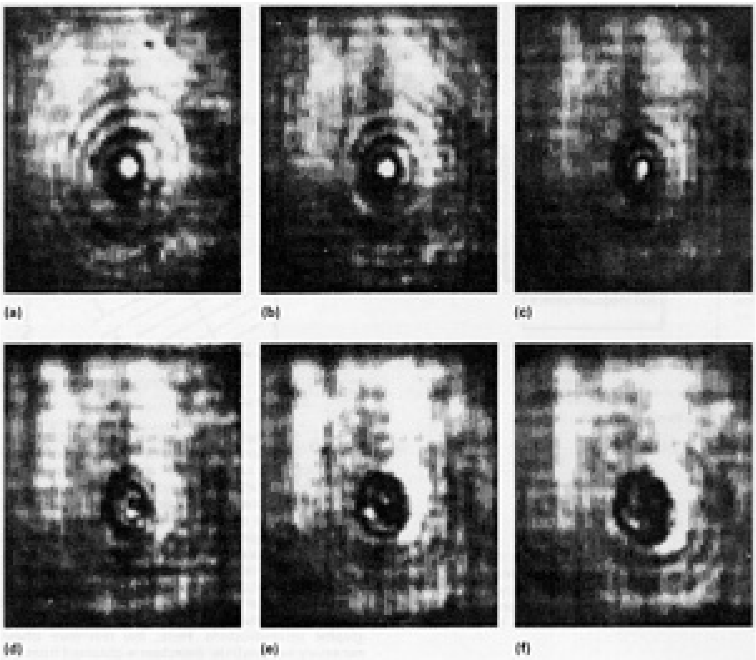


Fig. 11 SLAM image of an iron inclusion in an Si/SiC bending bar as obtained at the output detector. (a) z = 0 mm (the plane of detection at the surface of the sample). (b) to (f) Reconstructed images at various Δz (in increasing steps of 200 μm). As can be seen in (c) and (d), the defect appears focused, yielding a depth of approximately 500 μm. SLAM parameters: ultrasonic frequency, 100 MHz; field of view, 2.8 × 2.8 mm^2. Source: Ref 24

Summary

Acoustical imaging has gained tremendously from the comprehensive theoretical description of the contrast mechanisms involved and from the availability of high-speed computers. Such computers allow modeling and interpretation of the complex contrast underlying acoustic images and efficient handling of the large amount of data involved. Applications are primarily in NDE and materials characterization; some of these applications are related to problems in tribology. In the

future, high priority must be given to the integration of software that permits the reconstruction of defects by synthetic aperture techniques (Ref 2, 25, 26) and the use of robotic scanning systems in order to scan components of complex shape.

References

1. C.F. Quate, A. Atalar, and H.K. Wickramasinghe, Acoustical Microscopy With Mechanical Scanning--A Review, *Proc. IEEE,* Vol 67, 1979, p 1092-1113
2. A. Briggs, An Introduction to Acoustic Microscopy, *Microscopy Handbooks,* Vol 12, Oxford University Press, 1985
3. P. Höller and W. Arnold, Micro-Non-Destructive Testing of the Structure of New Materials, *Conference Proceedings of Ultrasonics International 1989,* Butterworth Scientific, 1989, p 880-888
4. L.W. Kessler and D.E. Yuhas, Acoustic Microscopy--1979, *Proc. IEEE,* Vol 67, 1979, p 526-536
5. R. Weglein, Acoustic Micro-Metrology, *IEEE Trans. Sonics Ultrasonics,* Vol SU-32, 1985, p 225-234
6. A. Atalar, Improvement of the Anisotropy Sensitivity in the Scanning Acoustic Microscope, *IEEE Trans. Ultrasonics, Ferroelectrics, Frequency Control,* Vol 36, 1989, p 164-273
7. J.I. Kushibiki and N. Chubachi, Material Characterization by Line-Focus Beam Acoustic Microscope, *IEEE Trans. Ultrasonics, Ferroelectrics, Frequency Control,* Vol SU-32, 1985, p 189-212
8. R. Weglein, SAW Dispersion in Diamond Films on Silicon by Acoustic Microscopy, *Rev. Quant. NDE,* 1992 (to be published)
9. A. Atalar, L. Degertekin, and H. Köymen, Acoustic Parameter Mapping of Layered Materials Using a Lamb's Wave Lens, *Proceedings of 19th International Symposium on Acoustical Imaging,* H. Ermert and H.P. Harjes, Ed., Plenum Press, 1992 (to be published)
10. J. Attal, L. Robert, G. Despaux, R. Capalin, and J.M. Saurel, New Developments in Scanning Acoustic Microscopy, *Proceedings of 19th International Symposium on Acoustical Imaging,* H. Ermert and H.P. Harjes, Ed., Plenum Press, 1992 (to be published)
11. A. Kulik, G. Gremaud, and S. Sathish, Direct Measurements of the SAW Velocity and Attenuation Using Continuous Wave Reflection Scanning Acoustic Microscope (SAMCRUW), *Acoust. Imaging,* Vol 18, 1991, p 227-236
12. K.K. Liang, S.D. Benett, B.T. Khuri-Yakub, and G.S. Kino, Precise Phase Measurements With the Acoustic Microscope, *IEEE Trans. Sonics Ultrasonics,* Vol SU-32, 1985, p 266-273
13. S.W. Meeks, D. Peter, D. Horne, K. Young, and V. Novotny, Microscopic Imaging of Residual Stress Using a Scanning Phase-Measuring Acoustic Microscope, *Appl. Phys. Lett.,* Vol 55, 1989, p 1835-1837
14. H. Vetters, E. Matthaei, A. Schulz, and P. Mayr, Scanning Acoustic Microscope Analysis for Testing Solid State Materials, *Mater. Sci. Eng.,* Vol A122, 1989, p 9-14
15. C.H. Chou and B.T. Khuri-Yakub, Acoustic Microscopy of Ceramic Bearing Balls, *Acoust. Imaging,* Vol 18, 1991, p 197-203
16. K. Yamanaka, Y. Enomoto, and Y. Tsuya, Acoustic Microscopy of Ceramic Surfaces, *IEEE Trans. Sonics Ultrasonics,* Vol SU-32, 1985, p 313-319
17. K. Yamanaka, Study of Fracture and Wear by Using Acoustic Microscopy, *Ultrasonic Spectroscopy and Its Applications to Materials Science,* Y. Wada, Ed., Special Reports of Japanese Ministry of Science, Education and Culture, 1988, p 44-49
18. S. Pangraz, E. Verlemann, and T. Holstein, unpublished results
19. I. Ishikawa, T. Semba, H. Kanda, K. Katakura, Y. Tani, and H. Sato, Experimental Observation of Plastic Deformation Areas, Using an Acoustic Microscope, *IEEE Trans. Ultrasonics, Ferroelectrics, Frequency Control,* Vol 36, 1989, p 274-279
20. I.N. Ermolov, The Reflection of Ultrasound From Targets of Simple Geometry, *Nondestr. Test.,* Vol 5, 1972, p 87-91
21. S. Pangraz, H. Simon, R. Herzer, and W. Arnold, Non-Destructive Evaluation of Engineering Ceramics by High-Frequency Acoustic Techniques, *Proceedings of the 18th International Symposium on Acoustical*

Imaging, G. Wade and H. Lee, Plenum Press, 1991, p 189-195
22. R.S. Gilmore, K.C. Tam, J.D. Young, and D.R. Howard, Acoustic Microscopy from 10 to 100 MHz for Industrial Applications, *Philos. Trans. R. Soc. (London),* Vol A320, 1986, p 215-235
23. Z. Lin, H. Lee, G. Wade, M.G. Oravecz, and L.W. Kessler, Holographic Image Reconstruction in Scanning Laser Acoustic Microscopy, *IEEE Trans. Ultrasonics, Ferroelectrics, Frequency Control,* Vol 34, 1987, p 293-300
24. A. Morsch and W. Arnold, Holographic Reconstruction by Back Propagation of Defect Images Obtained by Scanning Laser Acoustic Microscopy, *Proceedings of 12th World Conference on NDT,* J. Boogard and G.M. van Dijk, Ed., Elsevier Science, 1989, p 1617-1620
25. V. Schmitz, W. Müller, and G. Schäfer, Synthetic Aperture Focusing Technique--State of the Art, *Proceedings of 19th International Symposium on Acoustical Imaging,* H. Ermert and H.P. Harjes, Ed., Plenum Press, 1992 (to be published)
26. K.J. Langenberg, Applied Inverse Problems for Acoustic, Electromagnetic and Elastic Scattering, *Basic Methods of Tomography and Inverse Problems,* P.C. Sabatier, Ed., Adam Hilger, Bristol, 1987, p 125-467

Microindentation Hardness Testing

Peter J. Blau, Oak Ridge National Laboratory

Introduction

MICROINDENTATION (MICROHARDNESS) HARDNESS TESTING is an important tool for characterizing the near-surface characteristics of materials, surface treatments, and coatings. It is extensively used in both applied and research aspects of tribology. It is a subgroup of the general field of penetration hardness testing, but the relatively low applied forces (typically, 0.01 to 10 N, or 1 to 1000 gf) make it particularly sensitive to the near-surface mechanical properties of materials. Some common uses of microhardness testing include:

- Investigating the variations in penetration hardness between various phases in a microstructure
- Initial characterization of the surfaces of materials for wear applications
- Quality control of surface treatments or coatings
- Profiling the depth of hardened surface layers and coatings
- Assessing the nature of subsurface damage on or below machined surfaces
- Assessing the nature of subsurface damage on or below wear surfaces

In addition to hardness number determination, there are other specialized uses of microindentation techniques in friction and wear technology. These include:

- Use of microindentations as wear markers (see the section "Wear Measurement Using Microindentations" in this article)
- Use of microindentations to generate cracks to determine fracture toughness of brittle materials (Ref 1)

The development of instrumented microindentation testing equipment in recent years permits direct monitoring and recording of the instantaneous force versus the displacement depth for hardness tests. Such equipment offers the advantage of not requiring optical measurement of the indentation. However, it demands very accurate penetration depth calibrations. This type of testing is described in the article "Nanoindentation" in this Section. In this article, the focus will be on more traditional tests involving optical microscopy measurement of the impressions.

Principles of Microindentation Testing

The purpose of microindentation hardness testing is to obtain a numerical value that distinguishes between the relative ability of materials to resist controlled penetration by a specified type of indenter which is generally much harder than the material being tested. (A notable exception is in the microindentation testing of very hard materials, like diamond, where the indenter and test specimen can be equal or nearly equal in hardness.)

After preparing the specimen via the application of good metallographic practice in order to avoid residual damage to the test surface, the testing procedure involves the following sequence of steps:

1. Mounting the prepared specimen so that its test surface is perpendicular to the direction of indentation
2. Causing the indenter to move downward and impinge on the surface of the specimen at a specified rate
3. Allowing the indenter to remain for a specified residence time after it stops moving
4. Retracting the indenter
5. Measuring a characteristic dimension of the residual indentation
6. Using the geometry of the indenter to calculate a hardness number

Nearly all commercially available microindentation hardness testers perform steps 2 through 4 automatically.

The accuracy and reliability of the numbers obtained in performing microindentation hardness tests are strongly dependent on three factors: the machine, the operator, and the material characteristics. The machine must be correctly calibrated for both the applied force and the optical measuring accuracy. It must also be isolated from vibrations during the test. The operator must be familiar with the correct mounting and specimen preparation methods (such as rigidly mounting the specimen, keeping the test surface level, and using sound metallographic polishing practice to avoid the introduction of factors detrimental to specimen preparation), capable of measuring indentations consistently and correctly, able to recognize invalid indentations, and aware of the need to avoid touching the machine during its operation. The material may not be homogeneous or the method of fabrication applied in its production may give it hardness numbers significantly different from those published in tables of "typical values."

One of the greatest sources of error in determining microindentation hardness numbers is in the reading of the indentation lengths. This problem becomes particularly important when hard materials are being tested or low forces are being used.

Hardness numbers should not be operator dependent. Therefore, all the individuals using the given hardness tester should be tested to see how closely the measurements of indentation length agree on the same set of reference impressions. Personal correction factors may need to be given to each person so that measurements on reference specimens agree. Well-polished austenitic stainless steel or nickel specimens are good for laboratory optical reading reference specimens because they tend to provide nicely shaped impressions and remain untarnished. Periodic rechecking, approximately once a year, is desirable because an individual's vision is subject to change. If a critical series of measurements are to be made more frequently, then recalibration must be performed before each series is started.

If the hardness testing apparatus does not read directly in micrometers, each reader should be familiar with the proper eyepiece ("filar") unit-to-micrometer conversion method. A "filar" unit is a unit of measure that relates to a scale that is visible in the measuring eyepiece of the testing machine. If the filar eyepiece is used with different objective lenses, the conversion between the filar units and micrometers must be obtained for each objective lens. This conversion factor is derived by measuring the number of filar units that correspond to the observed spacings on a precision, etched microscope slide that is graduated directly in micrometers ("stage micrometer"). As noted above, each individual using the system should have his or her own personal filar factors to maintain consistency within the laboratory.

The numerical values obtained by microindentation hardness testing techniques are dependent on a combination of material properties (for example, elastic modulus, compressive yield strength, mechanical properties, anisotropy, and so on) that interact under the stress state imposed by the indenter. Therefore, hardness numbers should not be considered basic properties of a material or a coating, but rather numbers that indicate the response of a given material to the imposed conditions of the penetration test.

The two most commonly used microindentation techniques are the Vickers and the Knoop microindentation tests. Other indenter geometries have been developed for special purposes. These will not be discussed here; the reader is instead referred to books, published standards, and review articles included in Ref 2, 3, 4, 5, 6, 7, 8, 9, and 10.

Most commercial microindentation hardness testers still in use today use gages calibrated in gram-force (gf). However, the correct International Organization for Standardization (ISO) unit for force (that is, "load") is the Newton (N). Aside from the hardness scales that report a relative index or number of dimensionless units (such as Rockwell hardness numbers), indentation hardness numbers are commonly expressed in units of pressure. Units of force and pressure are related to the traditional microindentation hardness units as follows:

To convert from	To	Multiply by
gf	N	0.00981
kgf	N	9.81
kg/mm^2	GPa	0.00981
GPa	kg/mm^2	102.0

Historically, Knoop microindentation hardness is calculated as force per unit of projected area of the indentation, whereas the Vickers hardness (sometimes known as diamond pyramid hardness, or DPH) is expressed as a force per unit facet area. This difference in the area is one factor that can lead to slightly different hardness numbers under the same applied indenter force, especially at relatively low indenter loads (that is, 0.01 to 1 N, or 1 to 100 gf); Knoop numbers are usually higher than Vickers numbers obtained at the same load.

Standard Reference Materials for Microindentation Hardness

The U.S. National Institute of Standards and Technology, or NIST (formerly the National Bureau of Standards, or NBS), in Gaithersburg, MD, has developed materials whose microhardnesses can be used for calibration purposes. Known as Standard Reference Materials (SRM), they can be purchased for either Knoop testing or Vickers testing over the range of 0.25 to 0.98 N (25 to 100 gf). Each specimen comes with a certificate and a set of microhardness numbers for that specimen. Both materials are electroformed deposits and have the designations listed in Table 1.

Table 1 Standard reference materials used to calibrate Knoop and Vickers microindentation hardness equipment

Catalog No.	Scale	Material	Microhardness, kg/mm^2
SRM 1893	Knoop	Copper	125
SRM 1894	Vickers	Copper	125
SRM 1895	Knoop	Nickel	550
SRM 1896	Vickers	Nickel	550

Vickers Microindentation Hardness Test

The Vickers indenter is more widely used throughout the world than the Knoop indenter. The face-to-face angle of the Vickers indenter was selected so that Vickers hardness numbers would be comparable to those obtained from the established Brinell hardness test that preceded it. It was recognized that the diameter of Brinell hardness indentations (d_B) varied between 0.25 and 0.5 times the ball diameter (D_B). Using the mean of $d_B = 0.375\ D_B$ gave a ratio of surface contact area to the projected area (circular area of the indentation viewed from above) for the Brinell case of 1.08:1. This ratio is approximately the same for a Vickers pyramid when the face-to-face apex angle is 136°; hence, this angle was chosen for the Vickers pyramid. Figure 1 shows the shape of the tip of a Vickers hardness (diamond pyramid hardness) indenter and defines the symbols used in subsequent equations.

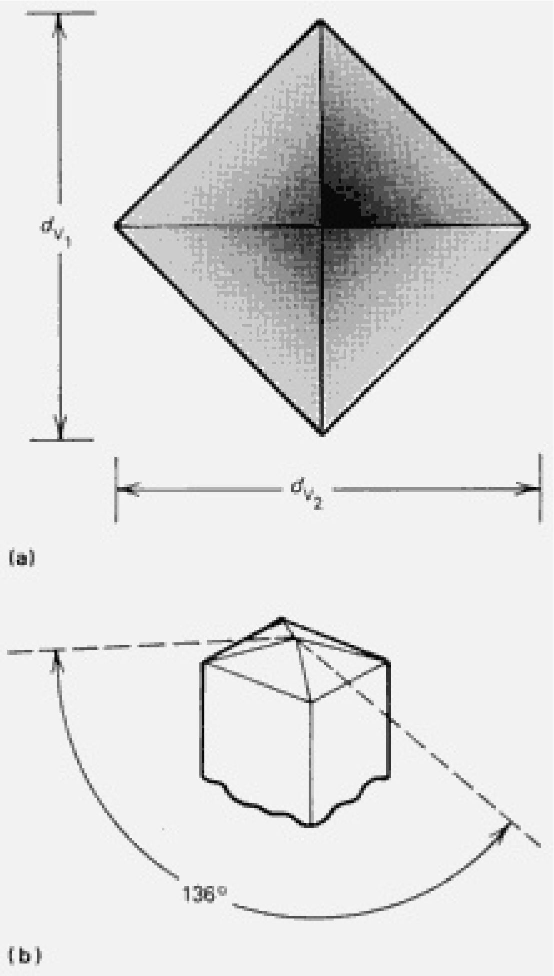


Fig. 1 Key dimensions and geometry for the tip of a Vickers indenter. (a) Diagonals d_{V_1} and d_{V_2}. (b) Face-to-face apex angle

There is some confusion in the literature as to the symbols used for microindentation hardness numbers using the Vickers indenter. Three commonly used symbols are DPH, VHN, and HV_P, where P represents the applied force. The later symbol is preferred by ASTM.

The general equation for Vickers hardness uses the average value of the two diagonals, *d**, where

$$d^* = \frac{(d_{V_1} + d_{V_2})}{2} \quad \text{(Eq 1)}$$

to calculate the hardness value

$$HV_P = C_v \left[\frac{P}{(d^*)^2} \right] \quad \text{(Eq 2)}$$

where P is the applied force, usually expressed in units of grams-force or kilograms-force, and C_V is a proportionality constant.

Vickers Indenter Versus Knoop Indenter. The Knoop indenter has both advantages and disadvantages over the Vickers indenter in microindentation testing. Because it is more blunt, the Knoop indenter tends to promote less cracking during the indentation of brittle materials (the penetration depth is 0.635 that of the Vickers indenter, assuming equal test

force and hardness number). Impressions are long and narrow, allowing the Knoop indenter to be used for hardness testing of thin layers or of narrow microstructural constituents. With the lower symmetry of the Knoop impression compared to the Vickers impressions, it tends to be somewhat more sensitive to crystallographic anisotropy than the Vickers indenter. The general shape of the ideal Knoop indenter tip and the symbols used in this section are shown in Fig. 2.

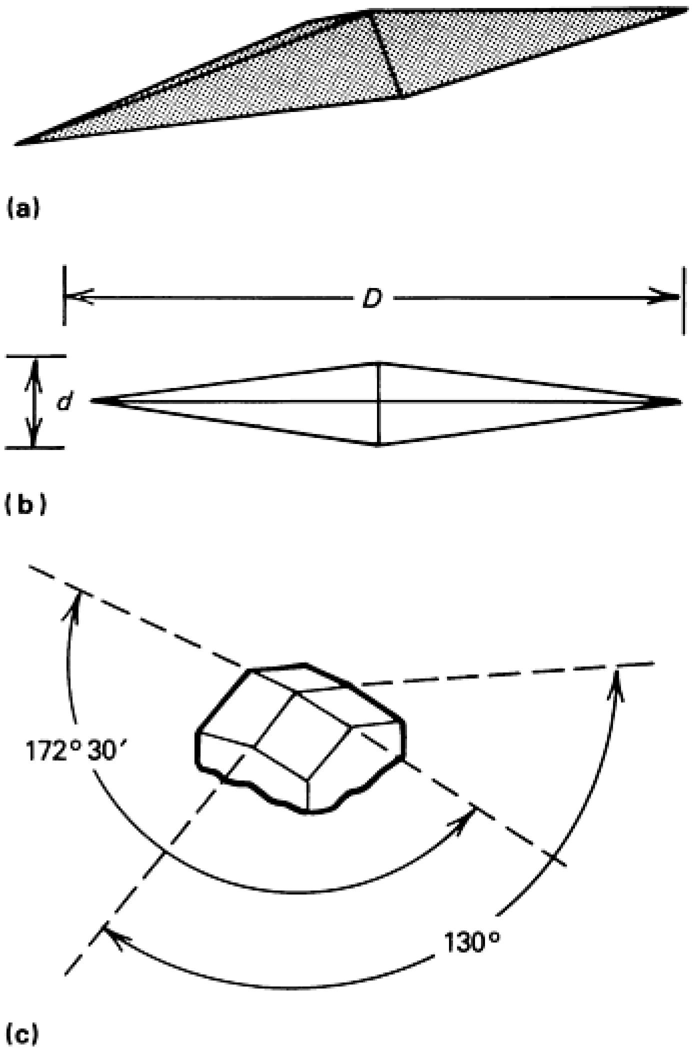


Fig. 2 Facets, dimensions, and geometry of a Knoop indenter tip. (a) Three-dimensional view of tip of Knoop indenter. (b) Diagonals *D* and *d,* where ratio of *D/d* is 7.1143. (c) Major (172.5°) and minor (130°) apex angles

Rather than requiring the averaging of two indentation diagonals, as in the case of the Vickers hardness calculation, the Knoop method requires only that the length of the longer diagonal, *D,* be measured. With the advantage of only one measurement per indentation comes the tendency for increased susceptibility to anisotropy in surface mechanical properties. The Knoop microindentation equation is

$$\mathrm{HK_P} = C_\mathrm{K}\left(\frac{P}{D^2}\right) \quad \textbf{(Eq 3)}$$

Table 2 gives the values for the proportionality constants C_V and C_K for various choices of units.

Table 2 Proportionality constants for Vickers and Knoop hardness scales as a function of hardness, force, and diagonal length

Constant		Applicable property value		
C_K	C_V	**Hardness**	**Force, *P***	**Diagonal length**
Knoop (HK)				
1.423×10^4	. . .	kg/mm^2	gf	*D*, μm
1.423×10^7	. . .	kg/mm^2	kgf	*D*, μm
1.396×10^2	. . .	GPa	gf	*D*, μm
1.396×10^5	. . .	GPa	kgf	*D*, μm
Vickers (HV)				
. . .	1.854×10^3	kg/mm^2	gf	*d**, μm
. . .	1.854×10^6	kg/mm^2	kgf	*d**, μm
. . .	1.819×10^1	GPa	gf	*d**, μm
. . .	1.819×10^4	GPa	kgf	*d**, μm

Microindentation Hardness Numbers of Materials

In the previous discussion, it was stated that microindentation hardness of materials depends not only on the composition of the material but also on the quality of the surface preparation, the method by which the material was produced, the indenter used, and the normal force (load) on the indenter. Therefore, the microindentation hardness numbers provided in this article are to be used only as a general guide to typical values for various materials. Owing to differences in surface preparation and microstructural condition, it is always better to obtain values on the specific specimen of interest than to use table values. Table 3 lists Vickers and Knoop microindentation hardness numbers that have been obtained on selected materials.

Table 3 Vickers and Knoop microindentation hardness numbers for selected materials

Material	Form[a]	Force, gf	Hardness numbers		Ref
			HV, GPa	**HK, GPa**	
Pure metals					
Aluminum	polyxl., 99.99%	10	0.25	. . .	11
	<100> on [100] xl. plane	50		0.227	12
Cadmium	High purity, polyxl.	25	0.21	0.21	13
		50	0.19	0.22	13
		100	0.19	0.23	13
Carbon	Diamond	200	98.65	. . .	14
Cobalt	High purity, polyxl.	25	2.46	2.88	13
		50	2.71	3.34	13
		100	2.59	2.73	13
Copper	High purity, polyxl.	25	0.57	0.48	13
		50	0.46	0.41	13
		100	0.42	0.37	13
Gold	Polyxl.	20	0.66	. . .	11
	. . .	10	. . .	0.77	15
Iron	High purity, polyxl.	25	2.44	2.92	13
		50	2.49	2.92	13
		100	2.44	2.76	13
Lead	Creeps at room temperature	5	0.049	. . .	1
Magnesium	Cast, 99.8-9% pure	40	0.47	. . .	11
Molybdenum	High purity, polyxl.	25	2.90	3.32	13
		50	2.73	2.76	13
		100	2.73	2.76	13
Nickel	High purity, polyxl.	25	1.00	1.06	13
		50	1.03	1.04	13
		100	1.01	1.14	13
Silicon	. . .	10	14.22	. . .	15
	Phase in aluminum alloy	50	. . .	8.83	15
	Phase in aluminum alloy	100	. . .	13.7	13

Silver	Coinage grade	20	0.94	. . .	11
Titanium	Annealed	60	1.54	. . .	11
Zinc	Single crystal (0001)	10	0.163	. . .	11
	Single crystal (1450)	10	0.323	. . .	11
Bearing steel					
AISI 52100	Ball	50	8.41	. . .	13
		100	8.41	9.80	13
Carbides and nitrides					
Boron carbide	BC	50	23.54	. . .	15
Chromium carbide	Cr_3C_2	. . .	19.12	. . .	15
Hafnium carbide	. . .	50	28.57	. . .	15
Silicon carbide	. . .	25	17.65	. . .	15
	(Green)	100	24.53		15
	Another source (gray)	100	38.25		15
Titanium carbide	. . .	. . .	31.38	. . .	11
		100	24.23	. . .	15
Tungsten carbide	WC	. . .	15.69	. . .	11
	WC	100	16.30	18.44	15
	WC_2	50	29.42	. . .	15
Titanium nitride	TiN	30	. . .	21.18	15
		100	. . .	17.36	15
Oxides					
Aluminum oxide	γ-Al_2O_3	100	24.5	. . .	11
	Sapphire	100	25.5	. . .	11
Magnesium oxide	MgO, single xl.	50	11.2	. . .	11
Silicon dioxide	SiO_2	. . .	11.8	. . .	11
Titanium dioxide	Rutile	25	9.8	. . .	11
Yttriuim oxide	Y_2O_3	. . .	6.86	. . .	11

Source: Ref 11, 12, 13, 14, 15

(a) xl., crystalline; polyxl., polycrystalline.

Microindentation Testing of Coatings

Film-covered surfaces and coated surfaces present special problems for micromechanical properties and hardness determination, especially if the properties of the thin layer(s) are to be separated from those of the substrate. Sometimes it is not necessary to separate the properties of the surface layer from those of the substrate if it is the materials system and not its constituent parts that is functionally important in the given application.

Single-Layer Coated Surfaces. Figure 3 indicates two of the four possible situations that can occur when indenting a surface covered by a film or a coating. In Fig. 3(a), the indenter moves directly through the coating layer without disturbing the coating of thickness *t*. In actual physical situations, it is unlikely that this situation will occur. The coating that originally occupied the volume now taken up by the indenter tip must go somewhere. If it behaves in a ductile manner, some of the coating material will be displaced to the side of the indenter and other portions may be drawn down into the indentation (Fig. 3b). Some materials that are normally considered to be brittle may behave in a ductile manner under the highly localized and confined compressive stresses beneath a sharp indenter tip.

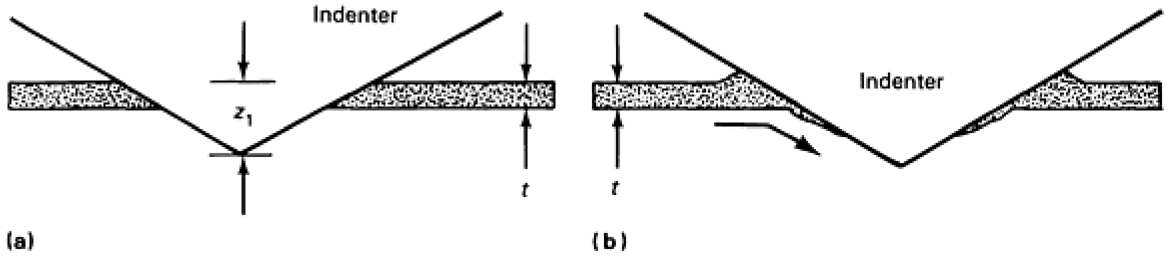


Fig. 3 Schematics showing two possible reactions of a coating layer to indentation. (a) In the ideal case, indenter moves directly through coating layer without disturbing coating thickness. (b) If coating reacts in a

ductile manner, the coating is drawn into the depression generated by the downward motion of the indenter tip.

The third possible reaction can occur if the coating contains significant levels of residual stress, if it adheres poorly to the substrate, if it is precracked, or if it is extremely brittle. That is, the coating will fracture or spall off in the neighborhood of the indentation. Trying to make sense of the hardness numbers obtained in this catastrophic coating failure is not worthwhile. In the fourth case, the coating may conform to the shape of the indenter and be pressed into the substrate without any change in thickness or the onset of fracture.

For the case where a portion of the coating material is drawn inward and another portion piles up to the side, the minimum facet contact area, *f,* occupied by the coating of thickness, *t,* when penetrated to a depth, z_1, can be determined by the equation:

$$f = 1 - \left[\frac{(z_1 - t)^2}{(z_1)^2}\right] \qquad \textbf{(Eq 4)}$$

From Eq 4, even when the penetration depth is twice the coating thickness, at least 75% of the indenter facet area is occupied by coating material. If the coating material is drawn into the indentation, this percentage may be higher, but the thin layer of coating material that coats the substrate at the bottom of the indentation is unlikely to have as important an effect on the load-bearing as the substrate itself.

Another issue involved with layered or coated surfaces is the strength of the bonding between the coating and the substrate. The action of indenting the surface may cause coating delamination to occur as the surrounding material buckles up and comes away from the substrate. Using this phenomenon, spherical indenters have been used in tests of polymer film bonding (Ref 16).

Several treatments for coating hardness have involved a rule of mixtures for the effective hardness number, H_{comp}. In one such treatment developed by P.J. Burnett and D.S. Rickerby (Ref 17), H_{comp} is determined to be:

$$H_{comp} = f_c H_c + f_s H_s X^3 \qquad \textbf{(Eq 5)}$$

where f_c and f_s are the respective volume fractions of coating and substrate materials being deformed; H_c and H_s are the respective coating and substrate hardnesses; and *X* is the interfacial parameter. The constraint parameters were found to be strongly dependent on the relative radii of the substrate and coating plastic zones. Also, for a weakly bonded film, the constraint parameter was closer to 1.0.

Jonsson and Hogmark (Ref 18) have developed an alternate formulation with the hardness of the film, H_f, given as:

$$H_f = H_s + \frac{H_c - H_s}{2Ct/D - C^2(t/D)^2} \qquad \textbf{(Eq 6)}$$

where H_s and H_c are the hardnesses of the substrate and the composite, respectively; *D* is the penetration depth (for Vickers, diagonal length/7); *t* is the film thickness; and *C* is the constant whose value depends on whether the film is brittle or whether it deforms to match the indenter shape. In the brittle case *C* is 0.0728, and in the ductile case *C* is 0.1403.

The use of two different treatments for the ductile and for the brittle films leads to the important issue of hard film on softer substrate versus soft film on harder substrate. The latter case is sometimes associated with the so-called "anvil effect." To obtain a "valid" hardness number for a material, ASTM standard E 384 requires that the penetration of the specimen material must be no more than one-tenth its thickness. This is a conservative position, because in some specific cases, the zone of affected material may extend to less than five times the thickness. In the absence of additional information about the material of interest, a factor of ten is generally reliable. If the penetration is not through the film or the coating, but represents say, one-half of its thickness, one would expect to see some effect on the hardness number by the substrate. Conversely, if the penetration depth is many times the film thickness, the effect of the film is expected to be

negligible. In one case where a submicron film of electrodeposited chromium on copper was Vickers hardness tested, it was found that the effect of the film was no longer significant when the penetration depth exceeded about twelve times the coating thickness.

When testing thin films or coatings, it is advisable to use a series of indentation loads. In doing so, a range of effects in the coating/substrate system can be examined. Empirical formulations can then be derived for comparing one coating system with another.

Multilayer coating systems (for example, electrical contacts) are even more complex to analyze mechanistically. The interaction of more than two constituents with different mechanical properties and the interfacial constraints between layers need to be considered. Scratch tests of various types are often used to assess the adhesion and durability properties of coatings. Scratch testing is a separate subject that is discussed in the article "Scratch Testing" in this Section.

References 19, 20, 21, 22, 23, 24, 25, and 26 may be helpful in regard to the microindentation testing of thin films and coatings. Some of these references use the nanoindentation techniques described elsewhere in this Volume.

Wear Measurement Using Microindentations

By utilizing the geometrical properties of microindentations, it is possible to estimate the amount of surface recession due to wear. For example, the ratio of the indentation depth to the length of the major diagonal of a Knoop indentation is 1:30.5, and the ratio of the depth of a Vickers indentation to the length of a diagonal is 1:7.00. These techniques can be applied only under the following two conditions:

- Wear surface material exhibits only a minor amount of elastic recovery after indentation, or the amount of elastic shape recovery is known
- Wear taking place is mild enough that the edges of the impressions are not unduly distorted or covered over

Microindentations can be used to measure mild wear in several ways:

- *Method 1:* periodical plastic replication of microindentations placed on the inside of a bushing or other normally inaccessible wear surface
- *Method 2:* calculation of incremental wear rate by measuring the changes in the dimensions of indentation diagonals
- *Method 3:* the smallest remaining indentation in a series that was originally produced at a succession of increasing loads

Method 1 (Ref 27). W.A. Glaeser used a Knoop indenter on the tip of a rigid rod tap a reference indentation inside a bronze bushing. By periodic replication of the indentation using plastic, Glaeser was able to watch the progression of wear in the bushing. In such cases, it would be advisable to place impressions at several locations because wear may vary from place to place on the contact surface of the bushing.

Method 2 makes use of the geometrical properties of indentations. When using this method, one assumes that the indentation geometry is similar to the indenter that produced the indentation. If substantial elastic shape recovery of the indentation occurs following the withdrawal of the indenter, then this method can only be considered to be approximate. By using the relationships between depth and diagonal length for Vickers and Knoop indenters, one can estimate a change in depth due to a change in the diagonal length due to wear:

$$z = \left(\frac{d_0 - d_f}{C}\right) \quad \textbf{(Eq 7)}$$

where z is the incremental wear depth, d_0 is the length of the diagonal of the hardness indentation before wear occurs, d_f is the length of the indentation diagonal after wear occurs, and C is a constant whose value is dependent on the type of

indenter used to generate the indentation. For a Vickers indentation with no significant elastic recovery in the surface, the value of C is 7.00; for a Knoop indentation with no significant elastic recovery in the surface, the value of C is 30.52.

This technique may be useful for measuring very small amounts of wear (for example, polishing). However, it is not effective when the wear process obscures the ends of the indentations. This method has been applied to determine wear patterns on magnetic recording heads (Ref 28).

Method 3 involves placing one or more rows of Knoop or Vickers indentations, produced at a series of loads, on the preworn surface of the specimen. After wear, the observer should note the smallest remaining indentation. Because a certain depth can be associated with each indentation, removal of a complete indentation indicates an increment of wear. If the hardness versus applied force behavior of the surface is known, indenter forces can be selected to give relatively equal depth increments. This wear measurement method has been suggested by L.K. Ives of the National Institute of Standards and Technology.

Correlation of Microindentation Hardness Numbers with Wear

There are many kinds of wear, as indicated in the Section titled "Wear" in this Volume. The types of localized stresses and directions of relative motion vary considerably between the various kinds of wear. Sometimes the stress conditions experienced by a wear surface are more similar to those imposed during a microindentation hardness test than in other cases. This means that the correlation between numerical values of microindentation hardness and wear rates may not necessarily correlate. Classic Russian studies of abrasive wear have established relationships of relative hardness to relative wear rates (Ref 29). However, this relationship does not always hold. There may not necessarily be a correlation between microindentation hardness numbers of the unworn surface and wear due to the following reasons:

Workhardening. In wear of metals and alloys, contact surfaces may change hardness greatly due to work hardening during wear. Therefore, using initial hardness numbers may be misleading.

Differences in Stress State. In a traditional hardness test, the indenter moves vertically down and up. In many forms of wear, material is deformed tangentially to the plane of the surface giving rise to shear stresses that do not occur in vertical penetration experiments.

Strain Rate Effects. In hardness testing, the rate of indentation (strain rate) may be small compared to that experienced by the same surface during wear. At high strain rates, materials are known to change their mechanical properties, often increasing in stiffness and yield strength.

Thermal and Chemical Contributions to Wear. Microindentation hardness numbers obtained under normal room environments may not accurately portray the mechanical behavior of a wear surface that is heated by friction. Furthermore, tribochemical effects such as oxidation or film formation may dominate wear behavior in ways that cannot be described by hardness numbers.

Third Bodies and Films. The microindentation hardnesses of the original surfaces may not bear any relationship to the properties of the material that accumulates on the wear surface during prolonged contact and that eventually governs wear behavior.

Effects of Brittleness. While many materials wear better if their surfaces are hardened, it is also true that very hard materials may be brittle and thus subject to fracture under the action of wear. The relationship of microindentation hardness to wear resistance may not be established for very brittle materials because the data simply cannot be obtained.

References

1. G.R. Anstis, P. Chantikul, B.R. Lawn, and D.B. Marshall, *J. Am. Ceram. Soc.,* Vol 64 (No. 9), 1981, p 533
2. D. Tabor, *The Hardness of Metals,* Clarendon Press, Oxford, 1951
3. B.W. Mott, *Microindentation Hardness Testing,* Butterworths, London, 1957
4. H. Bückle, Progress in Micro-Indentation Hardness Testing, *Met. Rev.,* Vol 4, 1959, p 49-100
5. D.R. Tate, A Comparison of Microhardness Indentation tests, *ASM Trans.,* Vol 35, 1945, p 374-389
6. J.H. Westbrook and H. Conrad, Ed., *The Science of Hardness Testing and Its Research Applications,*

American Society for Metals, 1973

7. P.J. Blau and B.R. Lawn, Ed., *Microindentation Techniques in Materials Science and Engineering,* STP 889, ASTM, 1985
8. V.E. Lysaught, *Indentation Hardness Testing,* Reinhold, 1949
9. "Test for Microhardness of Materials," E 384, *Annual Book of ASTM Standards,* ASTM
10. "Test for Microhardness of Electroplated Coatings," B 578, *Annual Book of ASTM Standards,* ASTM
11. A.A. Ivan'ko, *Handbook of Hardness Data,* U.S. Department of Commerce, National Technical Information Service, 1971, transl. from Russian
12. B.C. Wonsiewicz and G.Y. Chin, A Theory of Knoop Hardness Anisotropy, *The Science of Hardness Testing and Its Research Applications,* American Society for Metals, 1973
13. P.J. Blau, compiled from studies done at the National Bureau of Standards, 1980-1983
14. M.M. Khruschov and E.S. Berkovich, *Ind. Diamond Rev.,* Vol 11, 1951, p 42
15. B.W. Mott, *Micro-indentation Hardness Testing,* Appendix I, Butterworths Scientific Publishing, London, 1956
16. P.A. Engel and M.D. Derwin, *Microindentation Methods in Materials Science and Engineering,* STP 889, ASTM, 1985, p 272-285
17. P.J. Burnett and D.S. Rickerby, *Thin Solid Films,* Vol 154, 1987, p 403-416
18. B. Jönsson and S. Högmark, *Thin Solid Films,* Vol 114, 1984, p 257-269
19. M. Antler and M.H. Drozdowicz, Wear of Gold Electrodeposits: Effect of Substrate and of Nickel Underplate, *Bell Syst. Tech. J.,* Vol 58 (No. 2), 1979, p 323-349
20. E.H. Enberg, Testing Plating Hardness and Thickness Using a Microhardness Tester, *Met. Finish.,* Vol 66, 1968, p 48
21. M. Yanagisawa and Y. Motomura, An Ultramicro Indentation Hardness Tester and Its Application to Thin Films, *Lubr. Eng.,* Vol 43 (No. 1), 1987, p 52-56
22. P.I. Wierenga and A.J.J. Franken, Ultramicroindentation Apparatus for the Mechanical Characterization of Thin Films, *J. Appl. Phys.,* Vol 55 (No. 12), 1984, p 4244-4248
23. M. Nishibori and K. Kinosita, A Vickers Type Ultra-Microhardness Tester for Thin Films, *Jpn. J. Appl. Phys.,* Vol 11, 1972, p 758
24. J.B. Pethica, R. Hutchings, and W.C. Oliver, Composition and Hardness Profiles in Ion Implanted Metals, *Nucl. Instrum. Methods,* Vol 209/210, 1983, p 995-1000
25. M. El-Shabasy, B. Szikora, G. Peto, J. Szabo, and K.L. Mettal, Investigation of Multilayer Systems by the Scratch Method, *Thin Solid Films,* Vol 109, 1983, p 127-136
26. V.C. George, A.K. Dua, and R.P. Agarwala, Microhardness Measurements on Boron Coatings, *Thin Solid Films,* Vol 152, 1987, p L131-L133
27. W.A. Glaeser, *Wear,* Vol 40, 1976, p 135-137
28. A. Begellinger and A.W.J. de Gee, *Wear,* Vol 43, 1977, p 259-261
29. M.M. Khruschov, *Proceedings of Conference on Lubrication and Wear,* Institute of Mechanical Engineers, 1957, p 655

Nanoindentation

H. M. Pollock, School of Physics and Materials, Lancaster University (England)

Introduction

ONE REASON for the need to devote a separate article to nanoindentation, as distinct from microindentation, is that the type of instrumentation and data processing needed for nanoindentation has evolved in a different way. Additionally, some interesting phenomena that are especially important at the submicron scale occur, such as time-dependent behavior and the effect of ion irradiation. This article describes how nanoindentation data are obtained and interpreted, and provides examples of measurements involving polishing wear and other surface effects, surface treatments and coatings, fine particles, and other technologically important topics.

Nanoindentation Defined. Indentation testing becomes nanoindentation when the size of the indent is too small to be accurately resolved by optical microscopy. The concern is not solely with hardness and elasticity, because this definitely can be widened to include other types of tests, such as creep and even friction or film-stress measurement, which can conveniently be carried out with a nanoindentation instrument. In practice, the term nanoindentation usually implies the *continuous recording* of the distance moved by the indenter (penetration depth) and of the load, as well as other variables (such as time or frictional force), rather than single-valued measurements of contact area, as is usual with microindentation testing. This is simply because continuous depth recording (CDR) has proved to be the most direct solution to the problem of measuring very small indent sizes. However, as summarized in Table 1, not all nanoindentation testing involves CDR, which has its disadvantages. Conversely, CDR has proved to be a valuable addition to conventional microindentation instruments (Ref 1).

Table 1 Correlation (in practice) between CDR and micro/nanoindentation

Technique	Microindentation	Nanoindentation
Imaging of indents by microscopy	Usual	One type of commercial instrument only (Ref 2)
Continuous depth recording	Occasional examples in the scientific literature (see Ref 3)	Usual

The advantages of CDR include a high level of precision, ease of digitization, automation, and data processing. It is ideal for the measurement of creep, as well as of plastic and elastic work. Furthermore, each test normally gives a complete loading/unloading cycle, rather than just a single reading.

As a sole method of measuring indent size, its disadvantage is the need for simplifying assumptions in order to:

- Separate plastic from elastic effects
- Determine the true zero of the depth measurements
- Allow for piling-up or sinking-in of material around the incident
- Allow for geometric imperfection of the indenter when deriving absolute hardness values

Type of Information Obtained. Nanoindentation testing by CDR does not give values of absolute hardness directly. This is because hardness is usually defined as load divided by indent area projected onto the plane of the surface, and this area is not explicitly measured. However, the data can be processed on the basis of well-established assumptions (Ref 4) to yield relatively direct information that is of value in quality control.

This direct information on elastic recovery, relative hardness, work of indentation, and strain rate/stress relationship (Fig. 1) can provide a comprehensive "fingerprint" of a particular sample resulting, for example, from a change in either a production process or a wear test procedure. It is ideally suited to the comparison of one sample with a control or reference. The wider assumptions that are needed to derive indirect information on the material properties that are of particular scientific value are also defined in Fig. 1.

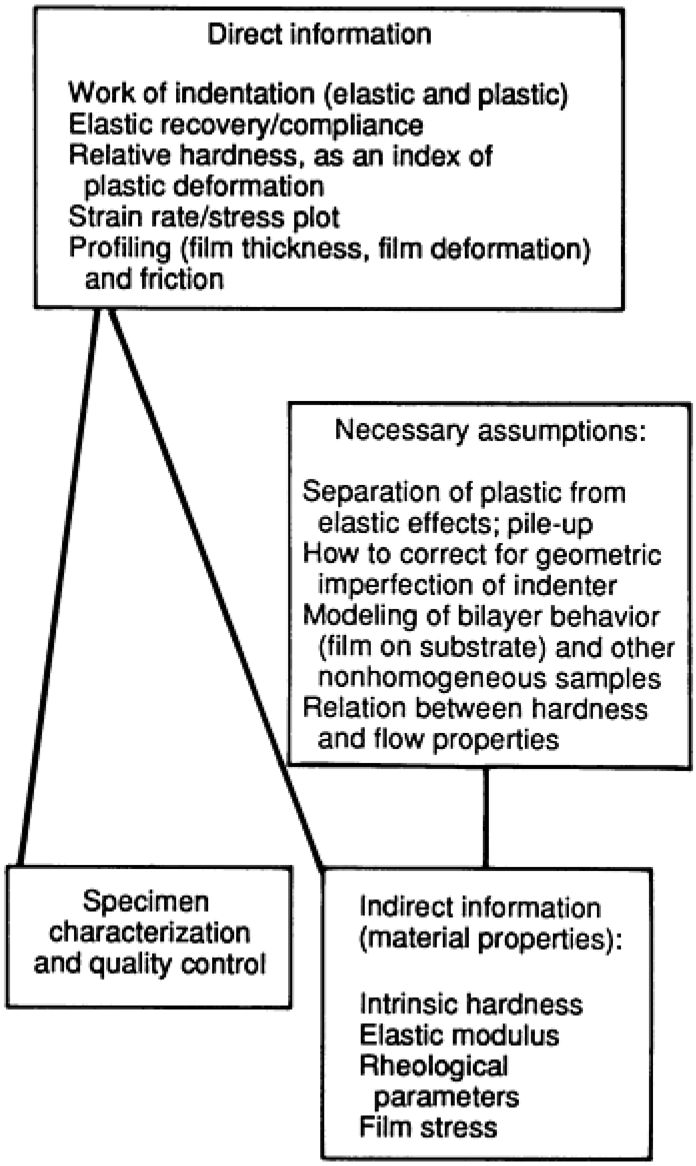


Fig. 1 Types of information obtained

The topic has produced some surprises, as mentioned in an earlier review (Ref 3) that discussed near-surface creep at low homologous temperatures, and some work in Japan that sometimes found a "critical load," below which no detectable plastic deformation remains. A number of years passed before independent confirmation was obtained, either experimentally (Ref 5, 6) or theoretically (Ref 7, 8).

The area under the depth-load curve is related to the work done by the indenter on the specimen. By subtracting the area under the unloading curve from the total area, W_T, the work, W_p, that is retained by the specimen (Fig. 5) is measured. For an elastic material, all the work is released upon unloading, that is, $W_p = 0$ and $\delta_p = 0$. For a plastic material, all the work is retained by the specimen, that is, $W_p = W_T$ and $\delta_p = \delta_T$. If the departure from linearity of the unloading curve is neglected, then:

$$\delta'_e = \frac{2W_e}{P_m} \qquad \textbf{(Eq 9a)}$$

and because R is defined as δ'_e/δ_p, with $\delta_p = \delta_T - \delta'_e$, Eq 2 in the preceding article follows directly. Alternatively, $P_m\, d\delta/dP$ can be substituted for δ'_e (Fig. 5), giving Eq 3.

To express R in terms of H and E, it is seen, from Eq 1, that $\delta'_e = (1 - \nu^2)P/(2Ea)$, and that from the definition of k_1, $\delta_p = a(\pi/k_1)^{1/2}$. Thus, $R = \delta'_e/\delta_p = P/(2Ea^2)(k_1/\pi)^{1/2}$. But, from the definition of hardness, $P/(\pi H)$ can be substituted for a^2, so that Eq 4 follows.

The quantity $H(1 - \nu^2)^2/E^2$, which is independent of k_1, can be calculated as follows: From Eq 4, $H(1 - \nu^2)^2/E^2 = 4R^2/(\pi k_1 H)$, and from the definition of k_1 and I_p, $k_1 H = I_p$, so that Eq 7 follows.

Acknowledgements

The way in which the nanoindentation test procedures are presented in this article reflects many discussions and much helpful criticism from Dr. R.H. Ion. I am grateful also to Dr. R.C. Rowe, Dr. J. Skinner, Dr. J.-C. Pivin, and Dr. M. Ghadiri for providing specimens for Fig. 3, 8, 10 and 12, and 15, respectively.

Nanoindentation Instruments

Where the prime requirement is to obtain absolute values of hardness in the sense of resistance to plastic deformation, a logical approach is to replace the optical microscope of a microindenter by an electron microscope. A nanoindentation attachment that can be used inside a scanning electron microscope (SEM), has formed the basis of patents and is commercially available (Ref 2). In principle, this approach makes it possible to establish a reliable comparison between nanoindentation hardness values and established scales of hardness numbers, such as those defined in national standards specifications. It is necessary to overcome the difficulties of imaging small indentations with sufficient contrast, and, at the smallest depths, to correct for the deformation of the required conductive layer of soft metal (Ref 9). In practice, many investigators have found the advantages of continuous depth recording to be overriding. Therefore, the discussion that follows deals primarily with nanoindentation testing using CDR.

The main features of existing instruments are listed in Table 2. Except when adhesion is being measured, the indenter is virtually always pyramidal, rather than hemispherical, given that the interest is in deformation at very small depths below the specimen surface. Vickers or Knoop indenters are sometimes used (Ref 1, 10), but a three-sided pyramid is more common, because this shape can achieve a better approximation to a perfectly pointed apex, either by polishing or by ion erosion. The apical angles can be chosen so that the nominal relation between indent area and depth is the same as for the Vickers shape.

Table 2 Potential design features of comprehensive nanoindentation instrument

Basic requirements
Vibration isolation and draft exclusion Indenter: normally a trigonal diamond pyramid
Soft testing designs
Loading device, such as coil in magnetic field
Depth-sensing transducer, such as capacitative, inductive, or fiber-optical
Analog or digital loading control (ramp mode; step mode)
Hard testing designs
Load cell (force-to-voltage transducer)
Displacement actuator
Analog or digital displacement control (ramp mode; step mode)
Data logging system
Data processing software
Options
Environmental enclosure
Optical microscope
Digital *x-y-z* sample displacement (from one indent position to the next; also, to microscopy location)
Additional ultrasmooth actuator for recalibration of depth transducer
Servocontrol of selected force or displacement
Modulation (ac component) of the load, for compliance measurements
Automatic detection of indenter-specimen contact
Additional programming: approach speed, loading/unloading cycle, series of indents
Vibration detector (data affected by vibration exceeding a set level to be discarded)
Temperature compensation
Hot stage

Friction measurement (smooth lateral displacement; transverse force transducer)
Scratch testing
Profilometry and depth sensing for measurement of film stress, Young's modulus, and other properties

Alternatively, a more acute angle (such as 90° between edges) can be chosen, on the grounds that thinner coatings can be tested without the data being significantly affected by the properties of the substrate. Also, sharp indenters give more consistent results when the specimen surface is rough (Ref 11). Like other mechanical instruments, nanoindentation devices can be either "soft" testing machines, where load is imposed and the displacement measured, or "hard" machines, where the displacement of the indenter into the specimen is imposed.

With soft machines (Ref 3, 12, 13, 14, 15), indenter and loading device must be mounted on a frictionless suspension. This often involves elastic design (spring hinges and leaf springs). In one instrument (Ref 13), an air bearing is used instead, and there is no need for the weight of the indenter to be counterbalanced. To minimize kinetic and impact effects, the moment of inertia of the moving assembly should be as small as possible. The relative movement of indenter and specimen (indentation depth) is measured by means of a displacement transducer, which can be a capacitance gage, a variable mutual inductance, or a fiber-optic device. The specifications of one commercial instrument (Fig. 2) are listed in Table 3.

Table 3 Commercial nanoindentation instrument specifications

Ultramicrohardness measurement	
Typical depth resolution	<1 nm
***X* resolution/travel**	0.02 μm/50 mm
***Y* resolution/travel**	0.02 μm/50 mm
***Z* resolution/travel**	0.02 μm/25 mm
Repositioning precision	0.1 μm
Typical force resolution	20 μN
Maximum load	0.5 N (0.11 lbf)
Software	Preprogrammed control and data analysis package with menu-driven facilities for custom program development
Dimensions	500 × 400 × 300 mm
Stress measurement	
Maximum sample size	200 mm diameter
Analysis field	50 × 50 mm square
Data obtained	Film thickness, film and substrate elastic modulus, substrate shape, and film stress
Curvature resolution	$(\Delta R)\infty, R$ corresponds to $R > 3 \times 10^4$ m
Friction measurement	
Maximum force	200 mN
Scan length	To 50 mm
Data obtained	Coefficient of static or dynamic friction

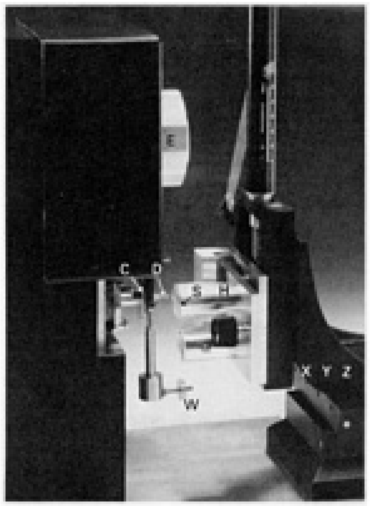


Fig. 2 Nanoindentation instrument with CDR; *xyz,* three-dimensional specimen micromanipulator; *H,* removable specimen holder; *S,* specimen; *D,* diamond indenter; *W,* balance weight for indenter assembly; *E,* electromagnet (load application); *C,* capacitor (depth transducer). Courtesy of Micro Materials Limited

With hard machines (Ref 10, 16, 17), the indentation depth is controlled, for example, by means of a piezoelectric actuator. Force transducers used in existing designs include: a load cell with a range from a few tens of μN to 2 N (Ref 17, 18); a digital electrobalance with a resolution of 0.1 μN, and a maximum of 0.3 N (Ref 16); and a linear spring whose extension is measured by polarization interferometry (Ref 10).

As noted in Table 2, it should be possible to vary the load or, in hard machines, the displacement, either in ramp mode or with a discontinuous increment (step mode). The important effects of varying the ramp speed, that is, the loading rate, will be discussed in the section "Choosing to Measure Deformation or Flow" in this article. The ramp function needs to be smooth, as well as linear, and there is evidence (Ref 19) that if the ramp is digitally controlled, the data will vary for the same mean loading rate according to the size of the digitally produced load increments, unless these are very small.

The basic requirements include a system for data logging and processing. Scatter in nanoindentation data tends to be greater than with microindentation, partly as a result of unavoidable surface roughness, but principally because the specimen volume being sampled in a single indentation is often small, compared with inhomogeneities in the specimen (such as grain size or mean separation between inclusions). Thus, unless such indent is to be located at a particular site, it is usually necessary to make perhaps five, ten, or more tests, and to average the data.

The spacing between indents must be large enough for each set of data to be unaffected by deformation resulting from nearby indents, and the total span should be at least one or two orders of magnitude greater than the size of the specimen inhomogeneities whose effect is to be minimized by averaging. On the other hand, if the test results turn out to be grouped in such a way that reveals differences between phases or grains, then each group should be averaged separately. In either case, the number of data points to be processed is large.

A real-time display helps the operator to monitor the data for consistency between indents and for any systematic trend and arises, for example, from a change in the effective geometry of the indenter, if traces of material from the specimen become transferred to it. The most common reason for an inconsistent set of data is a vibration transient, the effect of which is visible at the time. A subjective decision can then be made to discard that particular data set. Rather than use a real-time display for this purpose, a more reliable approach is to use the output signal from a stylus vibration monitor (a simple modification of the detection system itself) to abort any individual test during which the vibration exceeds a certain level.

Options that can greatly increase the scope and convenience of a nanoindentation instrument are listed in Table 2, in addition to basic requirements. With many specimen types, it is essential to record the exact location of each indent. This is achieved with the help of a specimen stage driven by either stepping motors or dc motors fitted with encoders. However, such devices must not be allowed to increase the total elastic compliance of the instrument to a value comparable with the smallest specimen compliance likely to be measured (the measurement of compliance is discussed in the section "Slow-Loading Test" of this article).

It is useful to be able to displace the specimen, as smoothly as possible, toward and away from the indenter, as well as to minimize impact effects at contact. This also facilitates recalibration of the displacement transducer, which in some designs varies according to the location of the plane of the specimen surface. In at least one design (Ref 3), the specimen displacement stage allows the surface to be brought into the field of view of an optical microscope, by using computer control. Another system (Ref 18) uses a closed-loop TV camera to help reposition the indenter rapidly and safely.

Refinements to the electronic hardware and software have been introduced to give, for instance, servocontrol of the selected load, loading rate, or displacement. This allows automatic compensation for nonideal transducer parameters, such as finite load cell compliance. Furthermore, the choice of loading mode (constant ramp speed or discontinuous step) can be extended to include more elaborate modes, such as constant strain rate or constant stress. Useful refinements include automatic control of the speed with which the specimen approaches the indenter and detection of the instant of contact. Thus, the whole loading-unloading cycle, and any required series of cycles, can be automated. As described in the section "Averaging of Multiple Tests" of this article, one design (Ref 6) includes provision for ac modulation of the load, which allows the continuous measurement of the compliance of the contact.

Of course, drift can be a problem, and attention must be paid to temperature stability. On occasion, temperature compensation is necessary in connection with depth or load transducers. Thus far, the introduction of specimen heating stages has been delayed by the consequent major problems of thermal drift.

Other physical measurements that require the use of the transducers mentioned also can be carried out, in principle, by means of a modified indentation procedure. One example is the determination of Young's modulus for thin films and other small specimens in the form of simple or composite beams whose elastic compliance is measured (Ref 16). As with optical scanning techniques, values of film stress can be derived from measurements of deflection and curvature of the film/substrate composite. Likewise, biaxial tensile testing of free-standing films can be carried out by means of the bulge test: if the bulge shape is profiled at a number of locations by probing with the indenter, then the strains can be calculated without the need to assume that the bulge is spherical. In effect, this represents a specialized type of profilometry, of which other examples include the measurement of film thickness and scratch width, as discussed below.

Film adhesion can be characterized by various methods, two of which can be used, in principle, with the help of a nanoindentation instrument modified to act as a film failure mechanism simulator. The indentation fracture technique (Ref 20, 21) has the advantage that normal loading only is required, thus avoiding complications of interpretation that arise from groove formation. In addition, values of fundamental parameters, such as critical stress-intensity factor, can be derived, in principle. A variant of the CDR technique, which monitors the load-depth curve, together with acoustic emission, in order to detect debonding at fiber-matrix interfaces in composites is described in Ref 22.

The thin-film scratch test has successfully been carried out by Wu *et al.* (Ref 23), who used conical indenters with hemispherical tips of radii down to l μm, and a tangential load cell. These were fitted to a nanoindentation instrument, for which the servo system could be set to give either constant indentation speed or constant rate of normal loading while the specimen was being translated at constant speed. As with conventional scratch testing, the critical load at which film cracking or delamination begins is used as an empirical measurement of adhesion. It was found that a reliable indication of this load was the value at which a load drop first occurred during a scratch loading curve. Thus, in many cases, fractography by SEM was not required for the detection of delamination.

Other established methods of detection, such as acoustic (Ref 24) or use of friction signal (Ref 25), can readily be used in conjunction with this technique. Wu *et al.* (Ref 18, 23) discuss the prospects of thus deriving values of fracture toughness of film/substrate assemblies. They also describe how scratch hardness is derived from measurement of the width of the scratch track when this is contained solely within one material (either bulk specimen or film only). The instrument could be operated in a simple profilometer mode, and values of track width were obtained from the observed difference between transverse depth profiles measured before and after the scratch was made.

Likewise, microfriction tests can readily be performed (Ref 26) if the indenter is replaced by the required friction stylus mounted on a device for measuring transverse force, such as a piezoresistive transducer. Again, the flat specimen is translated at constant speed. Simultaneous measurement of the stylus motion normal to the specimen surface, using the existing depth transducer of the basic nanoindentation instrument, when correlated with the peaks and troughs of the friction trace, can help to either confirm or eliminate different alternative models of the friction process (Ref 27). Nanoindentation has been used to characterize individual submicron-sized powder grains (Fig. 3), and the deformation and brittle fracture of spray-dried agglomerates has been recently quantified (Ref 28) with the help of an instrument modified by the addition of a crushing device.

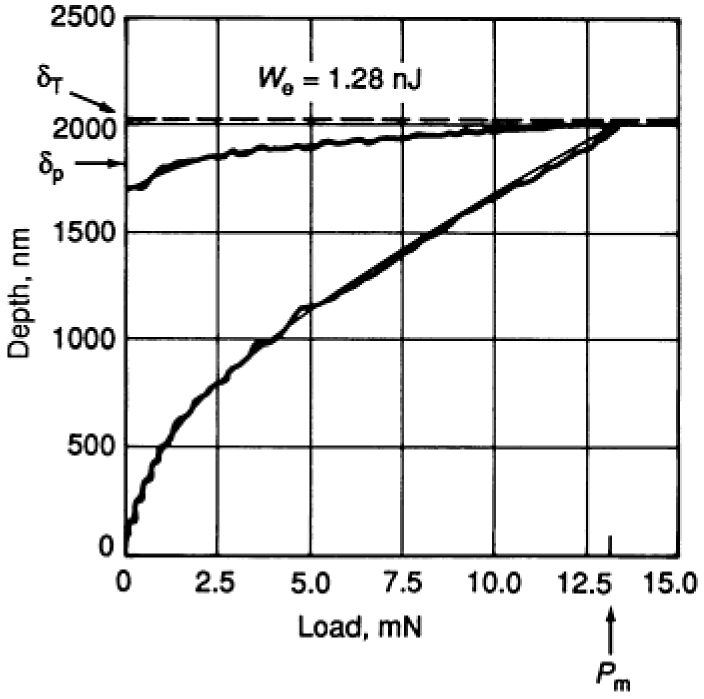


Fig. 3 Single test on individual 150 μm size grain of powder (lactose), with values of elastic recovery parameter, *R,* calculated by two methods: R_1 = δ'_e/δ_p = 0.095, and R_2 = $[(P_m\delta_T/2W_e) - 1]^{-1}$ = 0.107 (symbols defined in text and Fig. 5)

Test Procedures

Choosing to Measure Deformation or Flow. As yet, there is no universally accepted standard procedure or hardness scale that applies to nanoindentation with CDR. Consequently, the literature to date describes a variety of different data handling procedures, which generally have not yet been universally established. However, close examination shows that the differences are almost always a matter of presentation, rather than scientific content.

This review attempts to summarize all the principal techniques that have been published to date. Although the terminology used here has not necessarily been accepted in entirety, its usage is intended to emphasize the distinction made in Fig. 1 between the measurement of intrinsic material properties and the less ambitious task of characterizing particular specimens. Strictly speaking, terms such as loading rate will apply only when soft loading machines are used, but the equivalent hard loading procedure will be evident.

An assumption underlying the concept of hardness as a material property is that at or below some particular value of contact pressure, the plastic strain rate is zero. Unless the test is performed at the absolute zero of temperature, this is not strictly true. In many materials near the surface, indentation creep (including low-temperature plasticity) is often noticeable.

As discussed in an earlier review (Ref 3), if indentation depth varies significantly with timeas well as load, then even if the loading rate is held constant, and even if the material properties are independent of depth, there is no simple relation between load and depth. Furthermore, unless the indentation depth can be expressed in terms of separable functions of stress and time, the hardness, even if defined for a particular (constant) value of loading time or rate, will not be independent of load. Thus, as indicated in Fig. 4, a preliminary check is advisable.

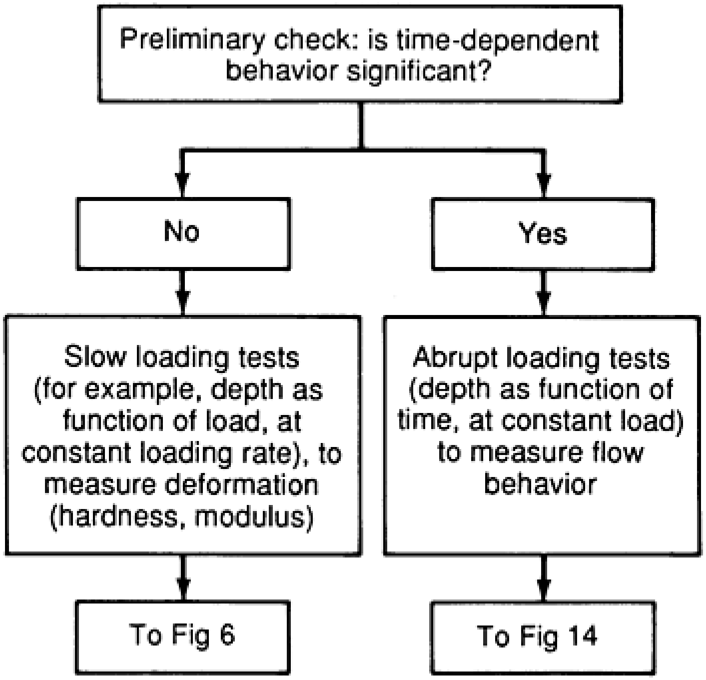


Fig. 4 Two principal types of test

The simplest way to measure deformation is by means of "slow-loading" tests, where the indentation depth is plotted as a function of slowly varying load, but it is wise to check the creep rate first by means of a load held constant for a time that is comparable to the duration of the proposed ramping load tests. Suppose that after that time, the creep rate is still *x* nm/s. Then, it would be reasonable to perform slow-loading tests in which the loading rate is always fast enough to produce a rate of indentation that is large, compared with *x*. If this is impractical, then rather than attempt a hardness test, it is logical to characterize the flow behavior, as discussed in the section "Flow Behavior" of this article.

Slow-Loading Test. Figure 5 shows a typical depth-load cycle, with load as the independent variable. Typically, a fresh location on the specimen surface is selected, and contact is made at a load of a few μN or less. The load is then raised at the required rate until the desired maximum is reached, and is then decreased, at the same rate, to zero. The "unloading curve," as shown, is not horizontal. The indenter is forced back as the specimen shows partial elastic recovery, and it is this phenomenon that allows the derivation of information on modulus. The amount of plastic deformation determines the residual, or "off-load" indentation depth, δ_p, and the plastic work, W_p (Fig. 5).

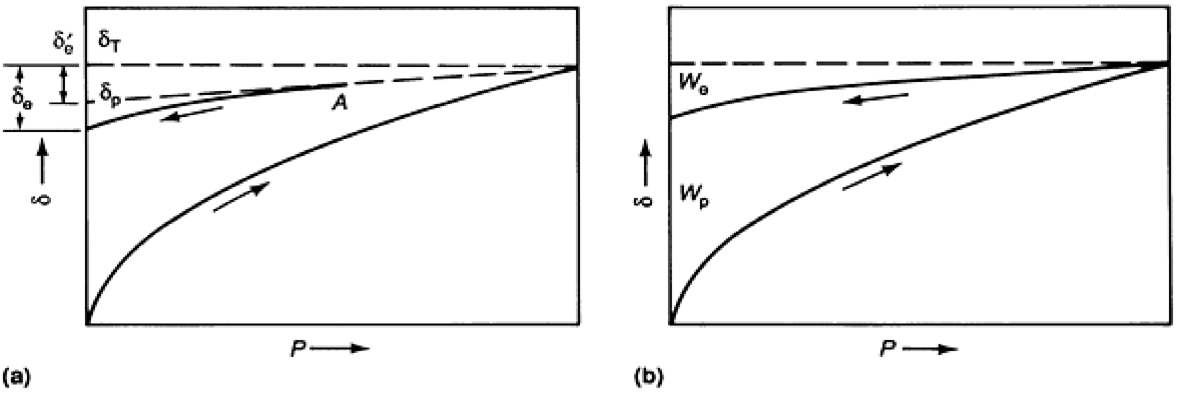


Fig. 5 Raw slow-loading data. (a) Depth, δ, as a function of load, *P*. (b) As (a), showing plastic and elastic work

A simple scheme for extracting information from such a test is shown in Fig. 6. Although there is no complete theory of elasto-plastic indentation, a useful approach is that of Loubet *et al.* (Ref 4). They used a simple approximation, namely that the total "on-load" elasto-plastic indentation depth, δ_T (Fig. 5), can be expressed as the sum of plastic and elastic components, δ_p and δ_e. It is further assumed that the area of contact between indenter ad specimen is determined by the plastic deformation only, and that δ_e represents the movement of this area as a result of elastic deformation (Fig. 7). If this were exact, δ_e would be given by Sneddon's relation (Ref 29) for a flat cylindrical punch normally loaded onto the plane surface of a smooth elastic body:

$$P = \frac{2Ea\delta_e}{1-\nu^2} \quad \textbf{(Eq 1)}$$

where P is the applied load, a is the radius of the contact region, E is Young's modulus, and v is Poisson's ratio. Thus, the unloading curve of δas a function of P would be linear.

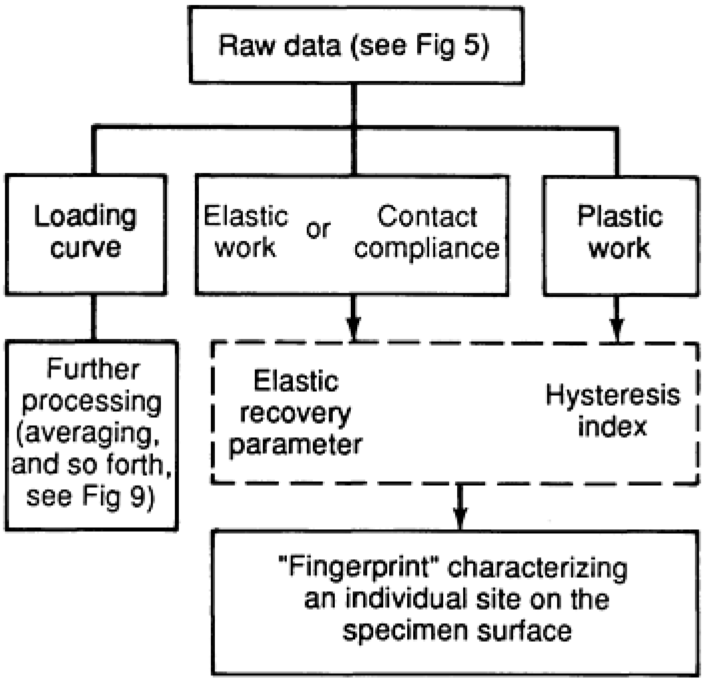


Fig. 6 Information from a single slow-loading test

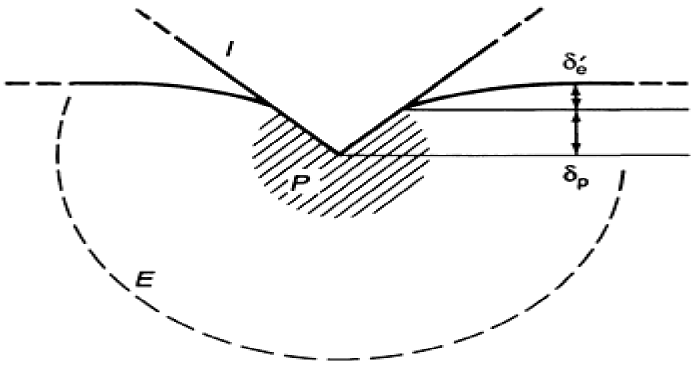


Fig. 7 Regions of elastic and plastic deformation (symbols as in Fig. 5); *I,* indenter; *P,* plastic zone; *E,* approximate limit of significant elastic deformation

In practice, there is some significant departure from linearity that occurs after a certain point (*A* in Fig. 5). This is attributed to a decrease in contact area arising from an opening of the apical angle of the indent, and a corrected value, δ'_e, is recommended instead of δ_e in Eq 1. Because δ'_e, as well as δ_T, can be determined experimentally, δ_p can be found. It is therefore possible to derive separate values of appropriate parameters describing the elastic and plastic behavior of the specimen.

Quite often, a typical specimen will show sizable variations in composition or structure, even within the small depth range sampled in these tests. Thus, before any attempt is made to derive values of material properties such as modulus, it is logical to define the most convenient indices that will provide a fingerprint characterizing an individual indent (Fig. 6). Ideally, these indices should relate directly to the raw test data, without the need for a sophisticated model or assumptions. In the simplest case of a homogeneous specimen whose material properties are constant, the values of these indices should also be constant, independent of depth. These conditions are satisfied by a fingerprint consisting of two numbers, an elastic recovery parameter, *R,* and a plastic hysteresis index, I_h.

The concept of an elastic recovery parameter was introduced by Lawn and Howes (Ref 30), who described its value in predicting how energy release provokes fracture in brittle materials. The definition was later modified (Ref 3, 31), so as to be consistent with the Loubet description above. It is defined as $R = \delta'_e/\delta_p$, and, as shown in Appendix , it is readily calculated from the area W_e (Fig. 5) that represents the mechanical work released during unloading:

$$R = \left(\frac{P_m \delta_T}{2W_e} - 1\right)^{-1} \quad \textbf{(Eq 2)}$$

where P_m is the maximum load. Alternatively, δ'_e can be found by fitting a tangent to the unloadingcurve at the maximum. In effect, this gives the contact compliance $d\delta/dP$, and thus (Appendix):

$$R = \left(\frac{\delta_T}{P_m \, d\delta/dP} - 1\right)^{-1} \quad \textbf{(Eq 3)}$$

Figure 3 shows an example. The physical significance of *R* is that for a homogeneous material, it is proportional to the ratio of hardness, *H,* to modulus $E/(1 - v^2)$, according to the formula (Appendix):

$$R = \frac{H(1 - \nu^2)(k_1 \pi)^{1/2}}{2E} \quad \textbf{(Eq 4)}$$

where k_1 is the geometrical factor that applies to the pyramidal geometry of the indenter used, namely the ratio of contact area (πa^2) to the square of the plastic depth (δ_P^2). The value *R* is a useful index, because it is a dimensionless quantity and because, in the case of an ideally homogeneous specimen, it will be independent of depth or load. Since *R* is derived from the contact compliance $d\delta/dP$ (according to Eq 3), any formula that includes *R* can of course be rewritten in terms of compliance.

The hardness will be proportional to P/δ_P^2, but a more direct measure of the resistance to plastic deformation is the hysteresis index, I_h, based on the plastic work, W_p. It is easy to show this in the simplest case of a specimen that shows fully plastic behavior with *H* independent of depth, $W_p \alpha \delta^3$ (Ref 3). Thus, I_h has the same dimensions as hardness if defined as:

$$I_h = P^3/W_P^2 \quad \textbf{(Eq 5)}$$

In the simple case mentioned, the value of *H* will be $I_h/(9\,k_1)$.

Figure 8 shows an example of indents located within individual grains in a two-phase material (cermet), illustrating the differences in *R* and I_h. It is important to realize that although I_h characterizes a plastically deformed zone whose

dimensions are not much greater than δ_p, the value of R is determined by the behavior of a much larger elastic hinterland (Fig. 7).

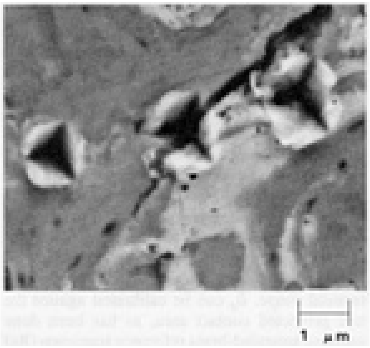


Fig. 8 SEM images of indents in a two-phase material (nichrome/chromium carbide cermet). Different maximum loads were used to give approximately the same indent depth in each case. Nanoindentation fingerprints, from left to right: Indent in dark region (carbide), R = 0.19, I_h = 754; mixed region, R = 0.19, I_h = 509; light region (nichrome), R = 0.12, I_h = 282

Averaging of Multiple Tests. In principle, a major advantage of CDR is the ability to obtain graphs of hardness as a function of depth from a single test. However, for the reasons outlined earlier, it is usually necessary to perform a number of tests on each specimen and to average the data (see Fig. 9). As it has been shown, each test yields only a single value of R, and it is advantageous to vary the maximum load between tests, so that R can be obtained as a function of depth. If the compliance of the instrument itself is significant, compared with the contact compliance $d\delta/dP$, it can be eliminated with the help of a plot of compliance against reciprocal of depth (Ref 32), extrapolated to infinite depth, before Eq 3 is used to calculate R. If $E/(1 - v^2)$ is known, the value of k_1 can be derived from a plot of this type (Ref 33).

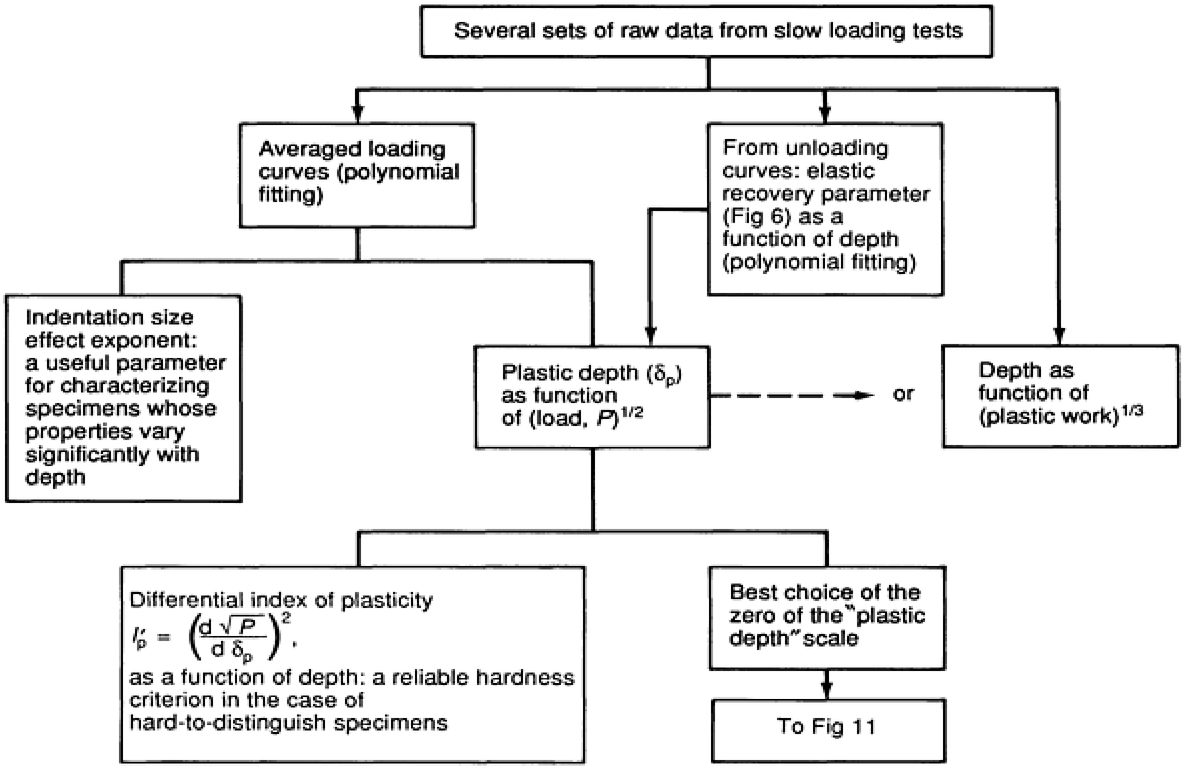


Fig. 9 Averaging of multiple tests

The need for discrete loading and unloading cycles to different values of maximum load is avoided if $d\delta/dP$ is continuously measured by means of a differential loading technique (Ref 6). An ac current source is used to add a small oscillatory modulation to the load, and the resulting oscillations in depth are detected with a lock-in amplifier. It is necessary to allow for the machine compliance and for damping inherent in the depth-sensing transducer.

No indenter pyramid is perfectly sharp, and with an indenter of finite tip curvature, plastic deformation is initiated at a finite depth below the surface. For this reason, and because of other complications (vibrational noise, pile-up around the indentation, and specimen roughness), the accuracy with which the instant of contact and the depth-zero can be identified is much poorer than the resolution of the depth measurement (typically, 1 nm or better). Methods that allow for the exact indent geometry are discussed in the section "Hardness and Modulus" of this article. In practice, the zero of plastic indentation depth can be determined with reasonable precision, for example, as shown in Fig. 9 and described below.

First, polynomials are fitted to the averaged loading curves and the graph of R against depth. From the relation $\delta_p = \delta_T - \delta'_e$, this allows δ_p to be plotted as a function of $P^{1/2}$. This function is chosen simply because, for an ideal material, it would be linear. In practice, after a certain depth has been exceeded, linearity is often seen over a considerable depth range, and the appropriate depth-zero can be found by extrapolation back to zero load (Fig. 10a). The result can be confirmed (Ref 3) by means of a (linear) plot of $W_p^{1/3}$ against depth, with extrapolation back to zero W_p.

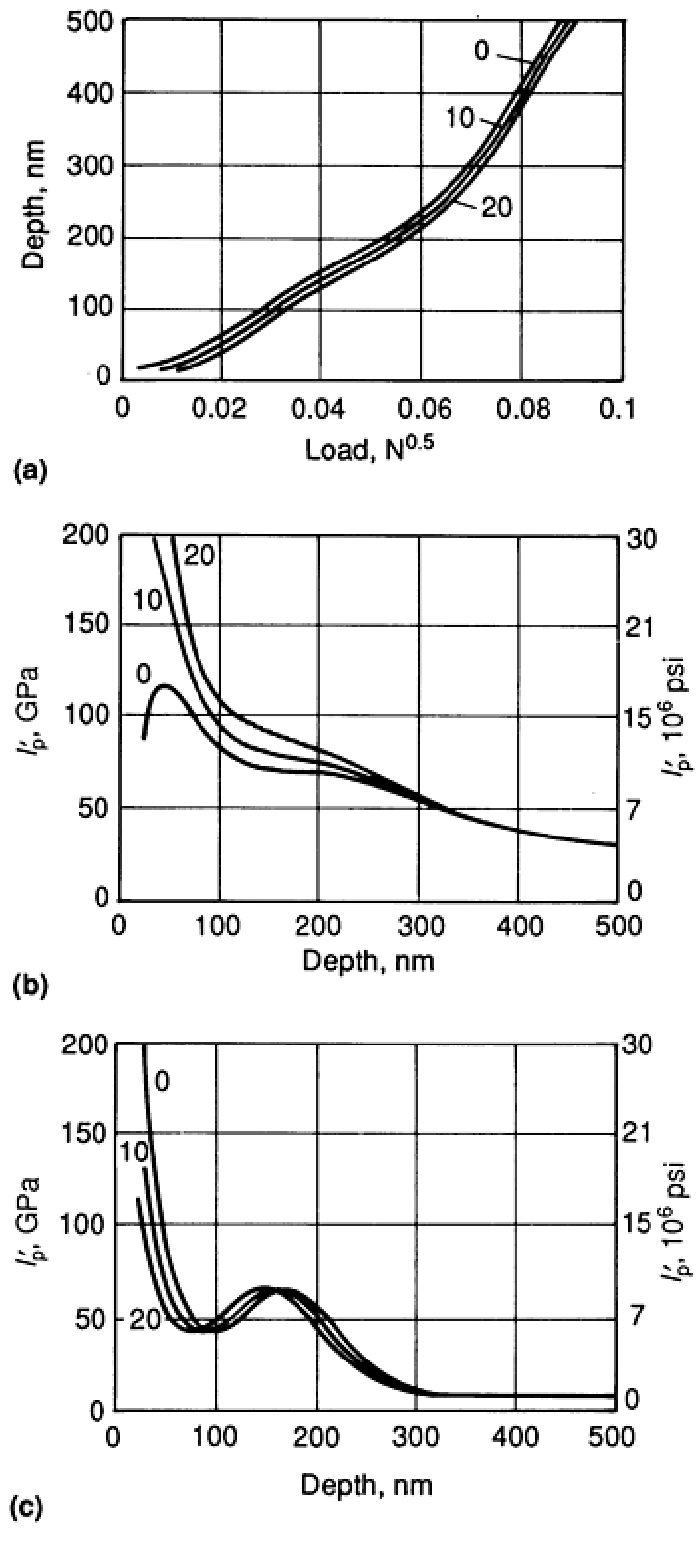


Fig. 10 Alternative data presentations, showing effect of changes in chosen depth-zero; numbers against curves indicate the depth offset in nm (specimen: multienergy boron implant into titanium). (a) Depth against square root of load. (b) I_p against depth. (c) I'_p against depth. Source: Ref 34

Subsequently derived profiles of relative hardness, in particular their shape at small depths, can show wild variations according to the choice of depth-zero (Fig. 10b), and should be interpreted with caution. For some materials (Ref 3), this near-surface difficulty is exaggerated by the interesting "critical load effect," whereby no permanent (plastic) indent size is made below a certain indent size. To study this, nanoindentation with depth recording was first used by Tazaki *et al.*

(Ref 35), whose work has been largely unacknowledged, although recently the effect has been confirmed for both sapphire and electropolished tungsten (Ref 6).

For hard ceramics, yielding can occur at intervals, with the observed increment in indent radius corresponding to the spacing (projected parallel to the surface) of discrete bands of deformation, separated by a characteristic distance (Ref 8); the relevant nanoindentation evidence is described in Ref 36. The critical load in, for example, sapphire or silicon carbide may correspond to the nucleation of the first plastic band. The formation of subsequent bands may be associated with the steps, or "serrated behavior" (Ref 17) often seen on loading curves.

Fortunately, when the aim is to characterize differences between samples, rather than absolute values of intrinsic hardness, the rather arbitrary effects of the choice of depth-zero may be minimized. Figure 10 illustrates one of the chief merits of nanoindentation with CDR, namely the fact that parameter values are derived for all depths up to the maximum reached in the series of tests. Because, in the case of an ideal material, hardness is directly related to the slope of the $\delta_p/\sqrt{P}$ curve, it is useful to convert this slope to give a parameter that has the dimensions of stress and to plot its value against depth, as shown. The definition of this differential index of plasticity (Ref 3) is:

$$\Gamma_p = (d\sqrt{P}/d\delta_p)^2 \quad \textbf{(Eq 6)}$$

As indicated in Fig. 9, this gives a reliable criterion for distinguishing between specimens of slightly different hardness, as well as for identifying small changes in penetration-resistance at certain depths. This is illustrated in Fig. 10(c), where the Γ_p curves show a feature at about 170 nm that is almost unnoticeable unless the data are presented in this way.

Hardness and Modulus. The idea of a differential hardness is physically obscure, and a measure of the hardness itself is often needed. As outlined in the second row of Fig. 11, this is possible subject to any uncertainties in the choice of depth-zero, and the comparative quantities listed are valid only when the same indenter is used for all tests. The bottom row of Fig. 11 lists quantities whose values can be used, in principle, when comparing specimens tested with different indenters. Here, the danger of systematic errors is, of course, greater.

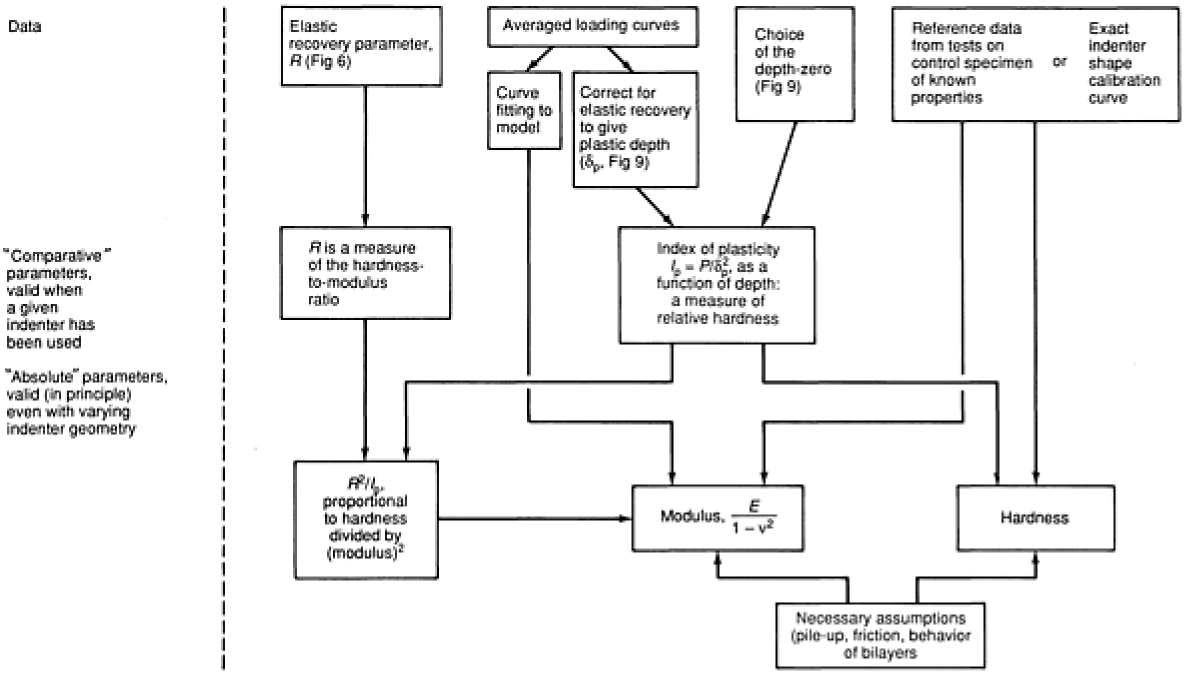


Fig. 11 Hardness and modulus

The comparative parameters include *R*, which is a measure of the ratio of hardness to modulus. As regards relative hardness, the appropriate index of plasticity, I_p, can be defined as P/δ_p^2 (Ref 3). Figure 12 shows examples of I_p as a

function of depth. For an "ideal" specimen and pyramidal indenter, the value of *H* will be I_p/k_1, but the danger of systematic error is reduced if values are normalized to data obtained from tests made on a control. These control data can themselves show a variation with depth. Accordingly, it can be useful to normalize each point on the graph to the individual value given by the control at that particular value of depth, rather than to a constant average value.

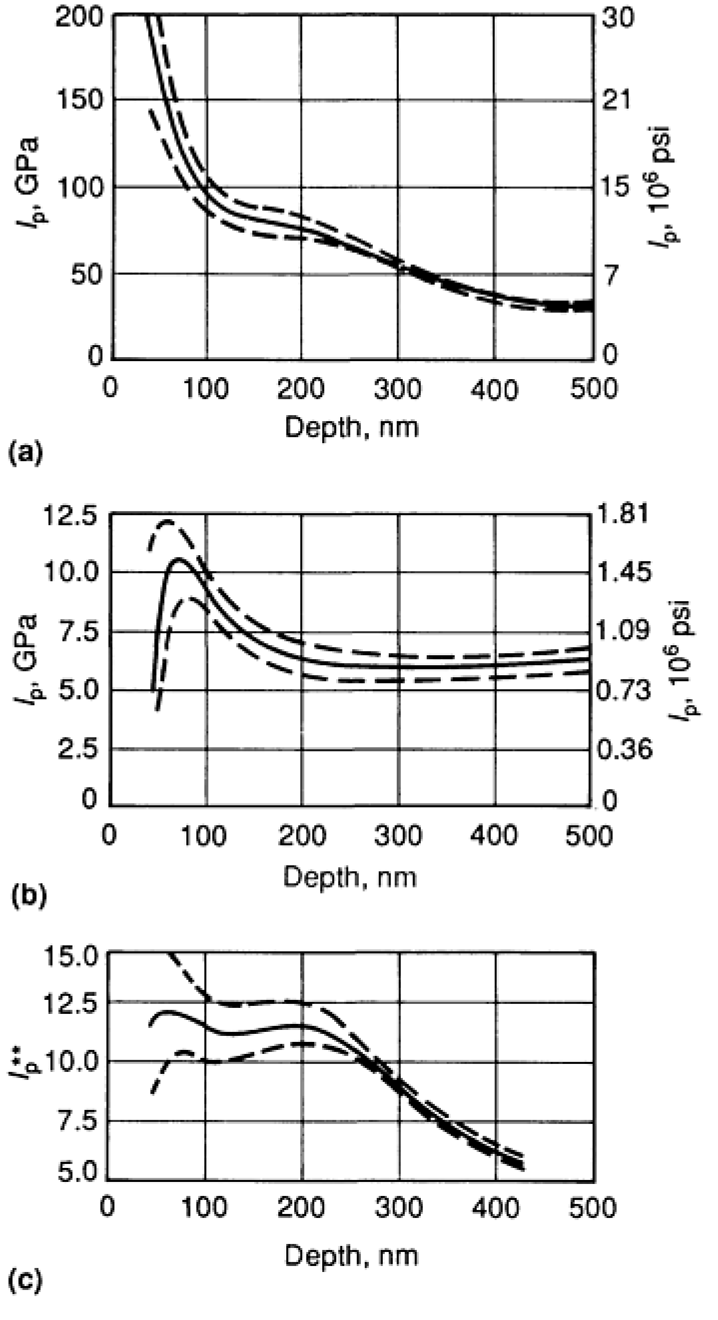


Fig. 12 Normalization at each value of depth. (a) Boron-implanted titanium, same as Fig. 10(b). (b) Nonimplanted titanium. (c) Curve a, normalized to curve b

In Fig. 12, the result is denoted by I^*_p. The parameter I_p or its equivalent has been used to quantify the relative hardness of a wide variety of surface layers, ranging from ion-implanted nickel (Ref 37, 38) and compositionally modulated multilayers (Ref 39) to titanium nitride layers produced by ion implantation (Ref 40) or reactive sputtering (Ref 41). C^+ implantation is known to increase the resistance of Ti_6Al_4V alloy to polishing wear, by a factor of nearly 100 (Ref 42). Unfortunately, the wear is not layer by layer, but is nonuniform over the surface, producing a rippled texture. Nanoindentation data were obtained and confirmed the reason for this behavior, namely that the polishing process produces its own strain-hardened gradient, with the more strained (hardened) regions wearing more slowly.

If the $\delta_p/\sqrt{P}$ curve is markedly nonlinear over the range of interest, it is clear that values of material parameters are varying significantly with depth. In this case, instead of plotting I_p, it may be simpler to derive the indentation size effect (ISE) exponent, *n*, which is related to the raw data more directly (see Fig. 9). The value *n* is defined by $P \propto d^n$, where *d* is the diameter or the depth of the pyramidal indent (Ref 43). (Sometimes *n* is termed the Meyer index, but it is important not to confuse it with the Meyer index that applies when a spherical indenter is used, and whose value depends on work-hardening behavior.) The *n* value is derived from log-log plots of δ_T versus *P* or, alternatively, from the formula (Ref 31):

$$n = P\delta_T/W_T - 1 \quad \textbf{(Eq 7)}$$

It has been used to detect the effect of traction (drawing) on the near-surface region of polymer films, as well as the point at which a substrate begins to influence the data obtained by testing a bilayer.

One precaution most needed when quantitative data are required is a regular check on the exact indenter geometry. As is evident from the above equations, any departure from the ideal pyramidal form, leading to an effective variation of k_1 with depth, or any change in its mean value, will alter the derived parameter values. This is particularly important if the aim is to derive absolute values, although as pointed out recently (Ref 44), the quantity $H(1 - \nu^2)^2/E^2$ is independent of k_1. As shown in Appendix , its value can be obtained from the formula:

$$\frac{H(1-\nu^2)^2}{E^2} = \frac{4R^2}{\pi I_p} \quad \textbf{(Eq 8)}$$

Clearly, the indenter geometry can be altered by accidental damage, but a more common cause is material that is transferred from specimens and adheres to the diamond. Clean, soft, work-hardening metals tend to be the worst culprits. Such contamination can often be removed by either ultrasonic cleaning or chemical attack.

A useful general cleaning procedure is to make a controlled indent into a polymer, such as polyethylene terephthalate (PET), in the hope that the contamination will adhere more strongly to the polymer than to the diamond. The regular checks on indenter geometry are typically performed before and after important data have been obtained, by simply making measurements on a control specimen of known properties. Suitable materials of reasonably uniform hardness include single-crystal silicon (001), for which no significant hardness anisotropy would be expected (Ref 45). With nanoindentation, experiment appears to confirm that the azimuthal orientation of the indenter is not important. Tungsten single crystals or thick, sputtered, pure metal films (Ref 17) have also been used.

Figure 11 emphasizes that in order to compare data obtained from tests made with different indenters, some type of indenter shape calibration is needed, because no indenter has a perfect pyramidal shape. δ_p can be calibrated against the true projected contact area, as has been done using an annealed-brass reference specimen (Ref 32). Two-stage carbon replicas of the indents were imaged in a transmission electron microscope. By means of this calibration, δ_p was converted into an "effective depth" that was exactly proportional to the square root of the area at all depths.

In effect, this procedure allows a constant value of k_1 to be used, so that *H* is given by I_p/k_1, and $E/(1 - \nu^2)$ is given by Eq 4. Time does not always permit the use of this delicate procedure, and a spherical cap model (a sphere of given radius capping a truncated pyramid) has been used (Ref 41) to given an approximation to the shape calibration function.

Whether an approach of this type is employed, or the less ambitious criteria discussed earlier (R^2/I_p; I^*_p) are used, certain factors that produce largely unpredictable variations in the resulting numerical data still remain. Several years ago (Ref 46), hardness values derived from nanoindentation at large depths were shown to agree to within 10 to 20% with values obtained from optical microscopy data, and agreement at this level should normally be expected, especially if k_1 is the same for the two indenters involved. However, the technique at present remains better suited to comparative testing than to the measurement of properties intrinsic to a given material, and no strictly valid correlation with Vickers microhardness values is possible (Ref 14).

Use of the above depth/area calibration procedures to derive so-called absolute hardness values neglects the fact that for most materials, even load divided by projected area will be different if a change from one indenter to another is made, because hardness varies with indenter shape. Moreover, the specimen under test and the reference specimen may have

significantly different characteristics with regard to indenter/specimen friction, as well as pile-up of material around the indentation, which is not detected in depth-sensing tests, but will affect the data (Ref 46).

Chaudhri and Winter (Ref 47) showed that for work-hardened mild steel and copper, the piled-up material supports a pressure equal to that supported by the rest of the indentation. As argued by Ion (Ref 34), the load-bearing area in such cases will correspond to a depth that exceeds δ_p, and may even be as large as δ_T. In other words, the on-load hardness can have more physical significance than the off-load hardness.

Apart from these factors, the main limitation on using the technique to measure material properties is the problem of interpreting data from specimens whose properties vary with depth, such as film/substrate bilayers. Of course, this is a problem at all levels of indentation testing, but is particularly important in nanoindentation because of the interest in indentation size effects, thin-film properties, and the often dominant effects of surface layers of oxide or contaminants.

For example, the presence of a very thin soft film can have surprisingly little effect on the raw loading curves (Ref 48). This is because at a given load, in the presence of the soft films (case A), δ_p will be greater than when no film is present (case B). In case A, however, the load is spread over a large area, so that δ'_e is less than in case B (see Eq 1, in which the elasticity of the substrate will be dominant). Thus, the difference between the totals $\delta_p + \delta'_e = \delta_T$ is quite small, and could be lost in experimental scatter.

When the elastic modulus of a film is calculated from elastic recovery or contact compliance data, the chief limitation is the fact that the effective value of *E* in Eq 4 is determined by the properties of a much larger volume of material (Fig. 7) than the plastically deformed region of hardness, *H*. An empirical model has been used (Ref 32) to describe the relative contributions of film and substrate to the measured compliance with appropriate weighting factors. Closed-form elasticity solutions have been used to model how the effective bilayer modulus varies with the projected area of the indent and agree well with experimental nanoindentation data (Ref 49, 50). The hardness of bilayers has been modeled by means of elastic-plastic finite-element analyses (Ref 51), and by an incremental kinematic method that takes account of frictional, as well as plastic, work (Ref 52).

In principle, it should be easier to measure film hardness if the film is softer than the substrate. This is because the value of mean pressure is affected more by the cumulative depth-integration of the work done by the indenter, than by whether or not the film has been penetrated. In the case of a film that is softer that than the substrate, even when the indenter (a 90° trigonal pyramid) has penetrated a distance of 1.4 times the film thickness, the measured hardness is still within 10% of the intrinsic hardness of the film (Ref 53). For films that are harder than the substrate, the critical depth, beyond which the influence of the substrate is significant, varies greatly with the indenter-specimen friction, and is less than the film thickness. Substrate, as well as film, deforms plastically at an early stage, and in cases of high friction, the film can be pushed down through a distance δ_p that exceeds the film thickness, without being pierced (Fig. 13). However, it still has proven to be possible to measure the intrinsic properties of a hard 300-nm thick Ni-25% B film on nickel, as shown by a plateau in the I_p versus δ_p graph (Ref 48).

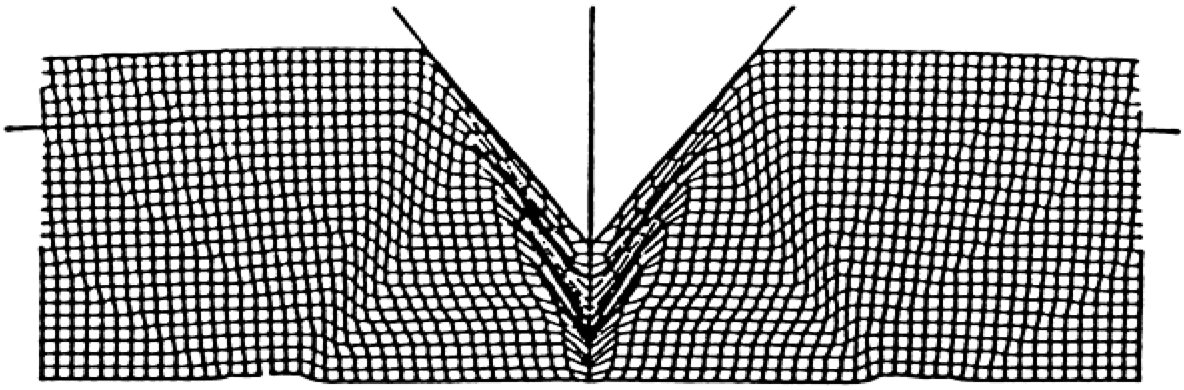


Fig. 13 Calculated field of deformation of hard film on substrate, with high friction. In this example, although the movement δ_p greatly exceeds the film thickness, the film has not yet been pierced (after Ref 53).

Flow Behavior. Exciting possibilities follow from the ability of the CDR technique to quantify the behavior of a specimen under load, even if the preliminary check (Fig. 4) suggests that slow-loading tests will be of little value. Recent calculations (Ref 7) show that most materials will exhibit indentation creep at temperatures as low as room temperature. The very high stresses involved induce dislocation glide as the principal mechanism, unless the grain size is less than a few hundred nm. The prediction has radical implications for the design and use of hard materials. Moreover, as argued in more detail elsewhere (Ref 3), the concept of a hardness that varies with time, with the theoretical complications involved, is not necessary. Instead, direct information on strain rate, as well as stress, and, hence, on the dislocation glide mechanism, is obtained without the need for a series of indentations with different loading times.

Consider an abrupt loading test, in which the load at the start is suddenly increased within less than a second to a chosen value, after which indentation depth is measured as a function of time. Subject to a number of simplifying approximations (Ref 3), the stress will decrease as P/δ_T^2, whereas the strain rate at any time is proportional to $\dot{\delta}/\delta$, that is, the speed of the indenter divided by the depth reached. (This last approximation follows from an earlier model, Ref 54, in which the strain rate is associated with the rate at which the plastic-elastic boundary of a spherical cavity moves on into the material.) Thus, this type of test corresponds to a vertical line on a Frost and Ashby temperature-stress deformation map (Ref 55), crossing successive strain-rate contours. The indenter never comes to rest (zero strain rate).

In principle, the flow mechanism can be identified, whether low-temperature plasticity (Ref 7), power-law creep as reported for stainless steel and aluminum (Ref 56), or various types of visco-plastic behavior, as shown by polymers (Ref 31), are involved. A major limitation is the difficulty of varying the specimen temperature without introducing thermal drift, the effect of which would dominate the observed value of $\dot{\delta}$.

The theory needed to interpret such data is at an early stage of development. If the main object is to characterize the difference in strain-rate sensitivity between specimens, then it is often wise to be satisfied with a graph of $\dot{\delta}/\delta$ against P/δ^2, on log-linear or log-log scales (see Fig. 14 and 15).

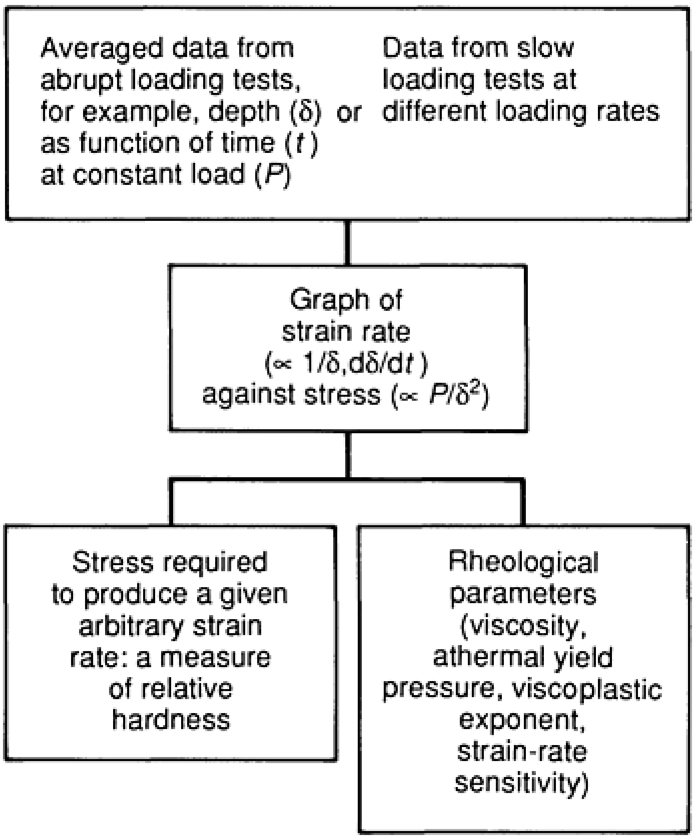


Fig. 14 Flow behavior

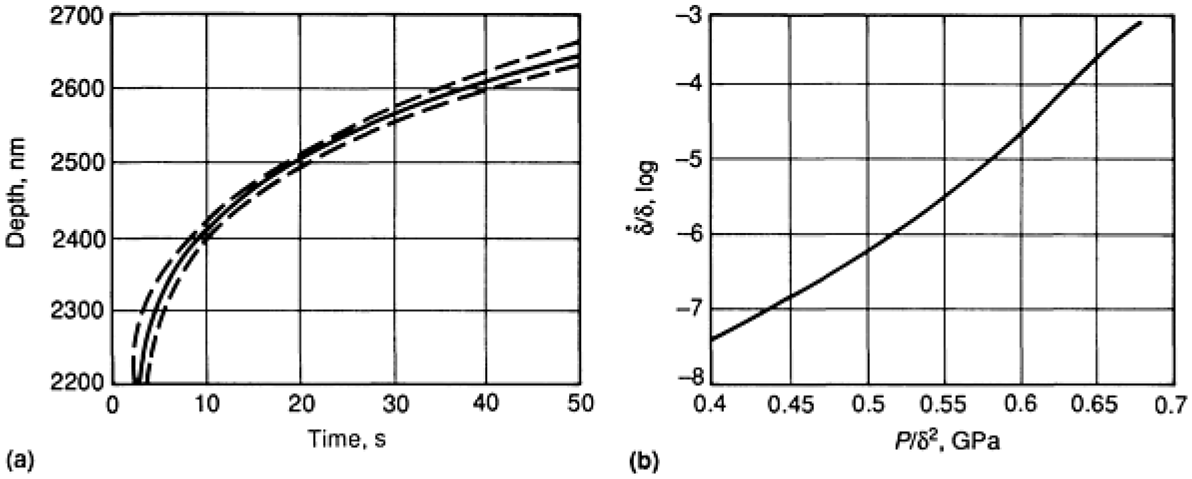


Fig. 15 Indentation creep of sodium chloride. Load, 3.3 mN. (a) Raw data. (b) $\dot{\delta}/\delta$ against P/δ^2

Variations from the above procedure have been described. With servocontrol, it should be possible to maintain $\dot{\delta}/\delta$ constant throughout a test or, alternatively, P/δ^2. Instead of these quantities being measured continuously (Ref 31), or over a small number of time intervals (Ref 57), conventional slow-loading (δ - P) tests at different loading rates have been performed (Ref 5) and analyzed to give information on strain-rate sensitivity.

Time-dependent, but recoverable (anelastic), deformation can be analyzed (Ref 31), following both the initial abrupt increase in load and a subsequent abrupt decrease in load to zero. After the initial plastic deformation appears to be complete, further changes in δ are dominated by recoverable behavior, with negligible further change in stress. Here, following the "flat punch" argument summarized earlier, the value of $\dot{\delta}$ can be taken as an indication of the anelastic strain rate.

Materials whose strain-rate behaviors have so far been studied include superplastic Sn-38% Pb, a nanophase ceramic (TiO_2) of a grain size from 5 to 12 nm, submicron films of Al on Si with both good and poor adhesion, and a polymer, PET. From log-log plots of stress $P/(k_1\delta^2)$ against strain rate $\dot{\delta}/\delta$, the strain-rate sensitivity, *m*, defined as d log(stress)/d log (strain rate), was measured for Sn-38% Pb, and the results supported a core-mantle model of superplastic deformation as an explanation of the observed enhancement of strain-rate sensitivity by grain boundaries (Ref 5).

Nanophase TiO_2 is more strain-rate sensitive than the single-crystal material (Ref 57), suggesting that the material in bulk form could show significant ductility. With a film-substrate system, the yield zone is confined by the interface (Ref 18), and for Al on Si, it was found that with good adhesion, the strain rate was lower at a given stress. Surface layers of PET show an anelastic compliance that fits well to a simple formula characterizing a lightly cross-linked polymer with a broad spectrum of relaxation times (Ref 31). The total recoverable deformation increases at high strain rates. Indentations to depths of less than 300 nm reveal that the near-surface layers are pseudoplastic, the behavior being further from ideal plasticity in the case of PET that has been uniaxially stretched or drawn.

Future Trends

Nanoindentation with continuous depth recording is being increasingly used in the characterization of submicron layers, surface treatments, and fine particles. Modified instruments are used also to measure film stress, thickness, adhesion, scratch hardness, and microfriction. Currently, the technique is best suited to providing a comprehensive quantitative fingerprint of the sample and to comparing it with a control, or reference. This direct information includes work of indentation, relative hardness, elastic compliance, and strain-rate/stress characteristics. There is no universally accepted absolute hardness scale that applies to nanoindentation. With the help of a number of assumptions, it is possible to derive values of intrinsic material properties, such as hardness or modulus, although it is not yet clear to what extent these values depend on test variables, such as the indenter geometry used. Furthermore, time-dependent behavior, for example, the effect of variations in loading rate, tends to be especially noticeable at the submicron scale.

A reading of the recent nanoindentation literature suggests that technical advances are likely to emphasize the following points:

- The processing of indentation creep data as an important aspect of material characterization
- The introduction of a new hardness scale (Ref 14) based on the method of continuous depth recording
- The development of multipurpose nanoindentation instruments that also perform scratch testing, profiling, and measurements of scratch hardness, film stress, friction, and other surface-mechanical properties
- Routine industrial testing, which requires improved automation so that the monitoring of changes in indenter shape, for example, is more reliably performed
- The introduction of specimen heating stages, together with a satisfactory method of compensating for thermal drift

In the longer term, picoindentation instruments are likely to be widely used to extend the technique to a still smaller scale, with the help of techniques developed for atomic force microscopy. Already, plastic deformation at depths of a few atomic layers, as well as the effect of surface forces, have been quantified by means of depth-load measurements, using a point force microscope, that is, an atomic force microscope operated in static (nonscanning) mode (Ref 58).

Appendix 1: Elastic Recovery Parameter

The area under the depth-load curve is related to the work done by the indenter on the specimen. By subtracting the area under the unloading curve from the total area, W_T, the work, W_p, that is retained by the specimen (Fig. 5) is measured. For an elastic material, all the work is released upon unloading, that is, $W_p = 0$ and $\delta_p = 0$. For a plastic material, all the work is retained by the specimen, that is, $W_p = W_T$ and $\delta_p = \delta_T$. If the departure from linearity of the unloading curve is neglected, then:

$$\delta'_e = \frac{2W_e}{P_m} \quad \textbf{(Eq 9a)}$$

and because R is defined as δ'_e/δ_p, with $\delta_p = \delta_T - \delta'_e$, Eq 2 in the preceding article follows directly. Alternatively, $P_m\, d\delta/dP$ can be substituted for δ'_e (Fig. 5), giving Eq 3.

To express R in terms of H and E, it is seen, from Eq 1, that $\delta'_e = (1 - v^2)P/(2Ea)$, and that from the definition of k_1, $\delta_p = a(\pi/k_1)^{1/2}$. Thus, $R = \delta'_e/\delta_p = P/(2Ea^2)(k_1/\pi)^{1/2}$. But, from the definition of hardness, $P/(\pi H)$ can be substituted for a^2, so that Eq 4 follows.

The quantity $H(1 - v^2)^2/E^2$, which is independent of k_1, can be calculated as follows: From Eq 4, $H(1 - v^2)^2/E^2 = 4R^2/(\pi k_1 H)$, and from the definition of k_1 and I_p, $k_1 H = I_p$, so that Eq 7 follows.

References

1. R.S. Polani, A.W. Ruff Jr., and E.P. Whitenton, A Dynamic Microindentation Apparatus for Materials Characterization, *J. Test. Eval.,* Vol 16, 1988, p 12-16
2. H. Baugert and A. Wagendristel, Ultra-Low Load Hardness Tester for Use in a Scanning Electron Microscope, *Rev. Sci. Instrum.,* Vol 56, 1985, p 1568-1572
3. H.M. Pollock, D. Maugis, and M. Barquins, Characterization of Sub-micrometre Layers by Indentation, *Microindentation Techniques in Materials Science and Engineering,* ASTM STP 889, P.J. Blau and B.R. Lawn, Ed., ASTM, 1986, p 47-71
4. J.L. Loubet, J.-M. Georges, O. Marchesini, and G. Meille, Vickers Indentation Curves of MgO, *Trans. ASME J. Tribol.,* Vol 106, 1984, p 43-48

5. M.J. Mayo and W.D. Nix, A Microindentation Study of Superplasticity in Pb, Sn and Sn-38 wt % Pb, *Acta Metall.,* Vol 36, 1988, p 2183-2192
6. J.B. Pethica and W.C. Oliver, Mechanical Properties of Nanometre Volumes of Material: Use of the Elastic Response of Small-Area Indentations, *Thin Films: Stresses and Mechanical Properties (Symp. Proc. 130),* J.C. Bravman, W.D. Nix, D.M. Barnett, and D.A. Smith, Ed., MRS, 1989, p 13-23
7. W.B. Li, K.E. Easterling, L. Henshall, and R.M. Hooper, The Mechanisms of Indentation Creep, *Acta Metall.,* Vol 39, 1991, to be published
8. S.J. Bull, T.F. Page, and E.H. Yoffe, An Explanation of the Indentation Size Effect in Ceramics, *Phil. Mag. Letters,* Vol 59, 1989, p 281-288
9. A. Wagendristel, H. Bangert, X. Cai, and A. Kaminitschek, Ultramicrohardness Measurements of Coated Samples, *Thin Solid Films,* Vol 154, 1987, p 199-206
10. B. Bhushan, V.S. Williams, and R.V. Shack, In-Situ Nanoindentation Hardness Apparatus for Mechanical Characterization of Extremely Thin Films, *Trans. ASME J. Tribol.,* Vol 110, 1988, p 563-571
11. D. Newey, H.M. Pollock, and M.A. Wilkins, The Ultra-Microhardness of Ion-Implanted Iron and Steel at Sub-Micron Depths and its Correlation with Wear-Resistance, *Ion Implantation into Metals,* V. Ashworth *et al.,* Ed., Pergamon, 1982, p 157-166
12. J.B. Pethica and W.C. Oliver, Ultra-Microhardness Tests on Ion-Implanted Metal Surfaces, *Ion Implantation into Metals,* V. Ashworth *et al.,* Pergamon, 1982, p 373-379
13. P.E. Wierenga and A.J.J. Franken, Ultra-Microindentation Apparatus for the Mechanical Characterization of Thin Films, *J. Appl. Phys.,* Vol 55, 1984, p 4244-4247
14. J.S. Field, Understanding the Penetration-Resistance of Modified Surface Layers, *J. Surf. Coat. Technol.,* Vol 36, 1988, p 817-827
15. C. Schmutz, J.P. Jeanneret, S. Tranganida, and H.E. Hintermann, Characterisation of Thin PVD Coatings by Microindentation, *Proc. Int. Conf. on Ion and Plasma-Assisted Techniques,* 1989, p 341
16. Y. Tusakamoto, H. Yamaguchi, and M. Yanagisawa, Mechanical Properties of Thin Films, *Thin Solid Films,* Vol 154, 1987, p 171-181
17. T.W. Wu, C. Hwang, J. Lo, and P. Alexopoulos, Microhardness and Microstructure of Ion Beam-Sputtered, Nitrogen-Doped NiFe Films, *Thin Solid Films,* Vol 166, 1988, p 299-308
18. T.W. Wu, Microscratch and Load Relaxation Tests for Ultra-Thin Films, *J. Mater. Res.,* Vol 6, 1991, p 407-426
19. M.J. Mayo and W.D. Nix, Measuring and Understanding Strain Rate-Sensitive Deformation with the Nanoindenter, *Strength of Metals and Alloys,* ICSMA 8, P.O. Kettunen, T.K. Lepisto, and M.E. Lehtonen, Ed., Pergamon, 1988, p 1415-1420
20. J.E. Ritter, T.J. Lardner, L. Rosenfeld, and M.R. Lin, Measurement of Adhesion of Thin Polymer Coatings by Indentation, *J. Appl. Phys.,* Vol 66, 1989, p 3626-3634
21. D. Stone, W.R. LaFontaine, P. Alexopoulos, T.W. Wu, and C.-Y. Li, An Investigation of Hardness and Adhesion of Sputter-Deposited Aluminium on Silicon by Using a Continuous Indentation Test, *J. Mater. Res.,* Vol 3, 1988, p 141-147
22. A.N. Netravali, D. Stone, S. Ruoff, and L.T.T. Topoleski, Continuous Microindenter Push-Through Technique for Measuring Interfacial Shear Strength of Fiber Composites, *Compos. Sci. Technol.,* Vol 34, 1989, p 289-303
23. T.W. Wu, R.A. Burn, M.M. Chen, and P.S. Alexopoulos, Microindentation and Micro-Scratch Tests on Sub-Micron Carbon Films, *Thin Films: Stresses and Mechanical Properties (Symp. Proc. 130),* J.C. Bravman, W.D. Nix, D.M. Barnett, and D.A. Smith, Ed., MRS, 1989, p 117-122
24. C. Julia-Schmutz and H.E. Hintermann, Micro-Scratch Testing to Characterize the Adhesion of Thin Layers, *Proc. 17th Int. Conf. on Metallurgical Coatings, Surf. Coat. Technol.,* Vol 48, 1991, p 1-6
25. J. Valli, A Review of Adhesion Test Methods for Thin Hard Coatings, *J. Vac. Sci. Technol.,* Vol A4, 1986, p 3007-3014
26. Q. Guo, J.D.J. Ross, and H.M. Pollock, The Role of Surface Forces in the Deformation and Adhesion of Solids, *New Materials Approaches to Tribology: Theory and Applications (Symp. Proc. 140),* L.E. Pope, L.

Fehrenbacher, and W.O. Winer, Ed., MRS, 1989, p 51-65

27. J.D.J. Ross, H.M. Pollock, and Q. Guo, Fine-Scale Adhesive and Frictional Interactions between Ceramics, *Powder Technol.,* Vol 65, 1991, p 21-35
28. K. Kendall and T.P. Weihs, Adhesion of Nanoparticles within Spray-Dried Agglomerates, *Proc. Conf. on Frontiers of Tribology,* Guildford, UK, 1991 (also to be published in *J. Phys. D: Appl. Phys.*)
29. I.N. Sneddon, *Int. J. Eng. Sci.,* Vol 3, 1965, p 47
30. B.R. Lawn and V.R. Howes, Elastic Recovery at Hardness Indentations, *J. Mater. Sci.,* Vol 16, 1981, p 2745-2752
31. R.H. Ion, H.M. Pollock, and C. Roques-Carmes, Micron-Scale Indentation of Amorphous and Drawn P.E.T. Surfaces, *J. Mater. Sci.,* Vol 25, 1990, p 1444-1454
32. M.F. Doerner and W.D. Nix, A Method for Interpreting the Data from Depth-Sensing Indentation Instruments, *J. Mater. Res.,* Vol 1, 1986, p 601-609
33. D.S. Stone, K.B. Yoder, and W.D. Sproul, Hardness and Elastic Modulus of TiN Based on Continuous Indentation Technique and New Correlation, *J. Vac. Sci. Technol.,* Vol A9, 1991, p 2543-2547
34. R.H. Ion, "An Evaluation of Ultra-Low Load Indentation Testing," Ph.D. dissertation, Lancaster University, England, June 1989
35. M. Tazaki, M. Nishibori, and K. Kinosita, Ultra-Microhardness of Vacuum-Deposited Films: II, *Thin Solid Films,* Vol 51, 1978, p 13-21
36. T.F. Page, W.C. Oliver, and C.J. McHargue, Deformation of Ceramic Crystals Subjected to Very Low Load (Nano) indentations, *J. Mater. Res.,* to be published
37. J.C. Pivin, J. Takadoum, J.D.J. Ross, and J.M. Pollock, Surface Hardness of Trinickel Boride--The Effect of Indent Size, Disorder, and Amorphization, *Tribology 50 Years On,* Mechanical Engineering Publications, London, 1987, p 179-181
38. D.S. Morrison, J.W. Jones, G.S. Was, A. Mashayekhi, and D.W. Hoffman, Characterization of Surface-Mechanical Properties and Residual Stresses in Ion-Implanted Nickel, *Thin Films: Stresses and Mechanical Properties (Symp. Proc. 130),* J.C. Bravman, W.D. Nix, D.M. Barnett, and D.A. Smith, Ed., MRS, 1989, p 53-58
39. R.C. Cammarata, J.E. Schlesinger, C. Kim, S.B. Qadri, and A.S. Edelstein, Nanoindentation Study of the Mechanical Properties of Copper-Nickel Multilayered Thin Films, *Appl. Phys. Lett.,* Vol 56, 1990, p 1862-1864
40. J.C. Pivin, F. Pons, J. Takadoum, H.M. Pollock, and G. Farges, Study of the Correlation between the Hardness and Structure of Nitrogen-Implanted Titanium Surfaces, *J. Mater. Sci.,* Vol 22, 1987, p 1087-1096
41. M.E. O'Hern, R.H. Parrish, and W.C. Oliver, Evaluation of Mechanical Properties of TiN Films by Ultra-Low Load Indentation, *Thin Solid Films,* Vol 181, 1989, p 357-363
42. I.L. Singer, R.N. Bolster, H.M. Pollock, and J.D.J. Ross, Polishing Wear Behaviour and Surface Hardness of Ion Beam-Modified Ti_6Al_4V, *Surf. Coat. Technol.,* Vol 36, 1988, p 531-540
43. P.M. Sargent, Indentation Size Effect and Strain-Hardening, *J. Mater. Sci. Lett.,* Vol 8, 1989, p 1139-1140
44. D.L. Joslin and W.C. Oliver, A New Method for Analysing Data from Continuous Depth-Sensing Microindentation Tests, *J. Mater. Res.,* Vol 5, p 1990, p 123-126
45. C.A. Brookes, J.B. O'Neill, and B.A.W. Redfern, Anisotropy in the Hardness of Single Crystals, *Proc. R. Soc. London,* A322, 1971, p 73-88
46. D. Newey, M.A. Wilkins, and H.M. Pollock, An Ultra-Low Load Penetration Hardness Tester, *J. Phys. E.,* Vol 15, 1982, p 119-122
47. M.M. Chaudhri and M. Winter, The Load-Bearing Area of a Hardness Indentation, *J. Phys. D.,* Vol 21, 1988, p 370
48. J.D.J. Ross, H.M. Pollock, J.C. Pivin, and J. Takadoum, Limits to the Hardness Testing of Films Thinner than 1 m, *Thin Solid Films,* Vol 148, 1987, p 171-180
49. D.S. Stone, T.W. Wu, P.S. Alexopoulos, and W.R. Lafontaine, Indentation Technique to Investigate Elastic

Moduli of Thin Films on Substrates, *Thin Films: Stresses and Mechanical Properties (Symp. Proc. 130),* J.C. Bravman, W.D. Nix, D.M. Barnett, and D.A. Smith, Ed., MRS, 1989, p 105-110

50. D.S. Stone, Elasticity Analysis to Aid in Extracting Thin Film Elastic Moduli from Continuous Indentation Data, *Trans. ASME J. of Electronic Packaging,* Vol 112, 1990, p 41-46
51. A.K. Bhattacharya and W.D. Nix, Analysis of Elastic and Plastic Deformation Associated with Indentation Testing of Thin Films on Substrates, *Int. J. Solids Structures,* Vol 24, 1988, p 1287-1298
52. D. Lebouvier, P. Gilormini, and E. Felder, A Kinematic Model for Plastic Indentation of a Bilayer, *Thin Solid Films,* Vol 172, 1989, p 227-239
53. J.C. Pivin, D. Lebouvier, H.M. Pollock, and E. Felder, Fields of Plastic Deformation in Indented Bilayers: Comparison between Kinematic Calculations and Experimental Data Obtained at Scales Ranging from 1 cm to 10 nm, *J. Phys. D.,* Vol 22, 1989, p 1443-1450
54. A.G. Atkins, A. Silverio, and D. Tabor, Indentation Creep, *J. Inst. Metals,* Vol 94, 1966, p 369-378
55. H.J. Frost and M.F. Ashby, *Deformation-Mechanism Maps,* Pergamon, 1982
56. S.-P. Mannula, D. Stone, and C.-Y. Li, Determination of Time-Dependent Plastic Properties of Metals by Indentation Load Relaxation Techniques, *Electronic Packaging Materials Science (Symp. Proc. 40),* E.A. Gies, K.-N. Tu, and R. Uhlmann, Ed., MRS, 1985, p 217-224
57. M.J. Mayo, R.W. Siegel, A. Narayanasamy, and W.D. Nix, Mechanical Properties of Nanophase TiO_2 as Determined by Nanoindentation, *J. Mater. Res.,* Vol 5, 1990, p 1073-1082
58. N.A. Burnham and R.J. Colton, Measuring the Nanomechanical Properties and Surface Forces of Materials Using an Atomic Force Microscope, *J. Vac. Sci. Technol.,* Vol A7, 1989, p 2906

Scratch Testing

Staffan Jacobsson, Michael Olsson, Per Hedenqvist, and Olof Vingsbo, School of Engineering, Uppsala University (Sweden)

Introduction

SCRATCH TESTING is a method of mechanically testing a specimen surface. In this method, a hard scratching element (indenter and stylus) is used to generate a groove in the specimen surface. The scratching action may, or may not, be accompanied by the formation of a chip or particles. The general objectives of performing scratch tests in materials research and testing are to:

- Clarify the mechanisms of deformation and/or material removal
- Evaluate or rank materials relative to abrasion resistance
- Measure scratch hardness
- Evaluate the adhesion of a surface coating to a substrate

The results of a scratch test can vary widely depending on the specimen material analyzed. Scratch test effects range from plastic grooving in a ductile material, to chipping in a brittle material, to interfacial deadhesion of a coated specimen.

Types of Scratch Test Devices

Apparatus Classification. Scratch test devices can be organized into three main categories (see Fig. 1):

- *Type 1:* low-speed bench top scratching machines, normally equipped with a stylus to produce a scratch on a flat with a single pass (Fig. 1a), with a reciprocating movement (Fig lb), or with a multiple pass (Fig. 1c)

- *Type 2:* low-speed scratching devices that operate *in situ* in the beam path of a scanning electron microscope (SEM); intended for detailed mechanism studies
- *Type 3:* high-speed scratching machines, which include penduli (Fig. 1d) (Ref 1, 2) and grinding wheels (Fig. 1e) (Ref 3, 4, 5); essentially used for single pass grooving

The categories can be further subdivided into machines for fixed-depth or fixed-load conditions. Type 1 and type 2 machines are usually equipped with deadweight or spring-controlled normal force, and the penetration depth is determined by the load and the deformation resistance of the workpiece material (that is, they are run under fixed-load conditions). In type 3 machines, which generate arc-shaped grooves, the normal force cannot be preset but is instead determined by the selected feed, the position along the groove, and the deformation resistance of the workpiece.

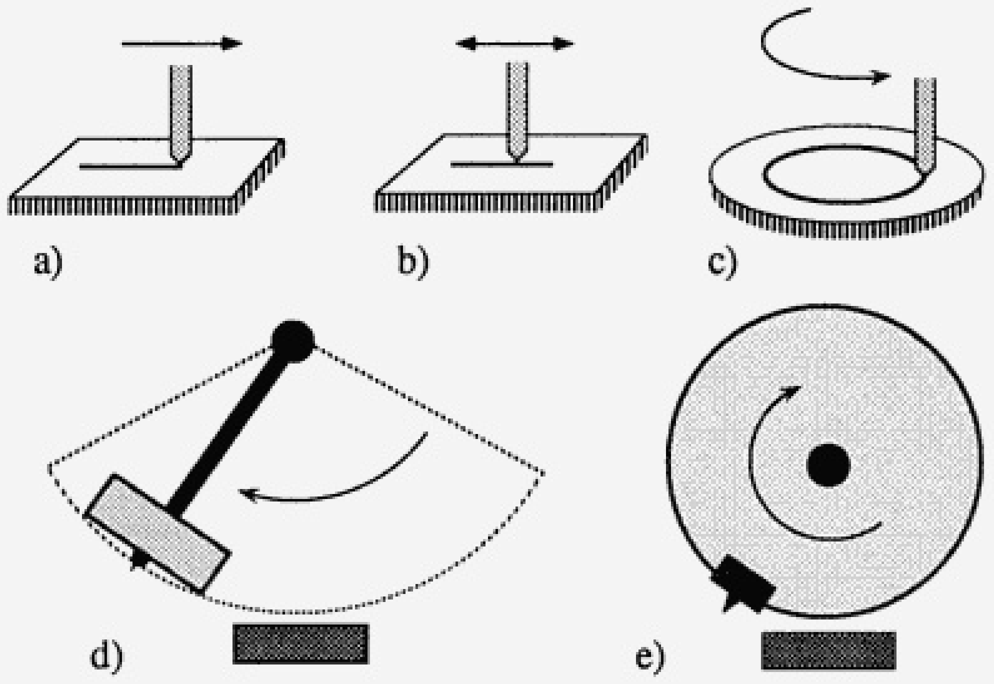


Fig. 1 Schematics showing setup and motion of components in selected scratch test devices. (a) Single pass. (b) Reciprocating. (c) Multiple pass. (d) Modified Charpy pendulum. (e) Modified grinding machine. Devices (a) to (c) are classified as low-speed machines. Devices (d) and (e) are classified as high-speed machines.

Measured Quantities. Many scratch testing devices are equipped with gages for continuous monitoring of the friction force. Furthermore, parameters such as the penetration depth and the acoustic emission are also sometimes monitored. Some methods involve posttest determination of groove shape, groove size, and specific grooving energy.

Scratching Elements. Scratch tip shapes can be classified as:

- Sharp conical
- Spherically rounded conical
- Sharp pyramidal
- Truncated pyramidal
- Irregular (abrasive particles)

Abrasive particles are employed in case studies of abrasive mechanisms, whereas the regular shapes are employed in investigations where maximum control and minimum geometric complexity are desired.

Because it is desirable that the scratching tip maintain its initial shape throughout a test series, it must be tough and substantially harder than the material to be scratched. The tip is ordinarily made of diamond, sapphire, tungsten, cemented carbide, or hardened steel. Rockwell and Vickers hardness test diamonds have well-defined shapes and are easily available. Thus, the Rockwell and Vickers hardness test diamonds have become the most commonly used scratching tools. Furthermore, it has been pointed out (Ref 6) that the width/depth ratio of a service groove typically varies between 5:1 and 50:1, a range that covers values for Vickers or Rockwell indenters.

***In Situ* Scratching Devices.** With the objective of producing detailed high-resolution studies of scratch formation mechanisms, a number of research groups have developed devices for *in situ* scratch testing applications in the beam path of a scanning electron microscope (Ref 7, 8, 9, 10, 11). The capabilities of such a device are shown in Fig. 2 (see the section "Scratch Adhesion Testing of Thin Hard Coatings" in this article). The limited volume of the specimen chamber sets limits on the size of the equipment, and the vacuum requirement of the SEM equipment restricts the use of any fluid lubricants or polymer specimens with a high degassing rate. Furthermore, the SEM technique requires electrically conductive specimens. Insulators can be studied if they are supplied with a conductive coating (for example, a sputtered gold film) prior to the investigation. However, this extra coating may affect the results of the analysis.

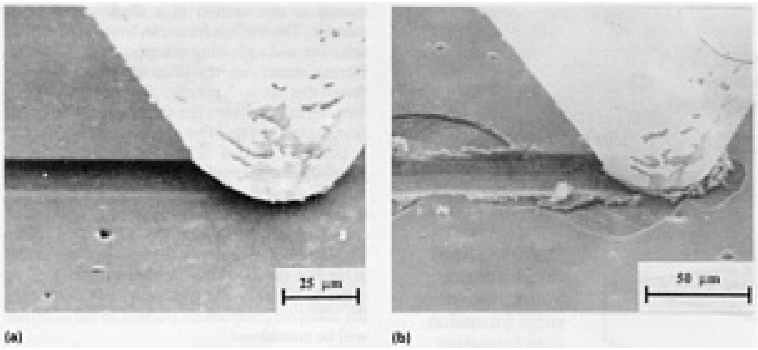


Fig. 2 Scratch channels from an *in situ* scratch testing experiment using an SEM apparatus. (a) $F_N < F_{N,C}$. (b) $F_N > F_{N,C}$. F_N, applied normal force; $F_{N,C}$, critical normal force

The observation of the deformation process with the SEM setup is facilitated if the scratching tip has a fixed position (typically centered on the monitor screen), while the specimen is slowly moved. Because the process is to be observed at high magnifications, the scratching speed has to be reduced to suitable values (for example, at 2000× magnification a speed of 5 $\mu m \cdot s^{-1}$ on the specimen table corresponds to a speed of 10 $mm \cdot s^{-1}$ on the screen).

Quick-Stop Devices. Quick-stop devices make it possible to abruptly stop the scratching action, thereby "freezing" the deformation process. The subsurface microstructure of the quick-stopped specimen is subsequently placed under a microscope to examine the polished and etched cross section of the groove. This technique facilitates highly detailed studies of the deformation mechanisms in front of and under the scratch tip (including the deformation of individual grains and phases, the formation of shear zones, cracks, and dead zones; see Fig. 3). Without the quick-stop technique, these studies would be impossible.

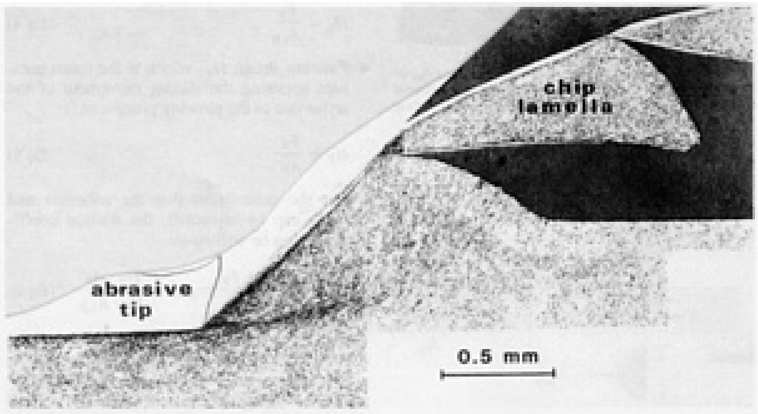


Fig. 3 Light optical micrograph of a cross section through a high-strength low-alloy (HSLA) steel specimen quick-stopped in the Uppsala pendulum. A truncated pyramid cemented carbide tip moving at a speed of 5.6 m/s (18 ft/s) was used as the scratching element.

The quick-stop event must be virtually instantaneous for the test to be relevant. Furthermore, the introduction of vibrations or extra movements must be avoided, or the microstructure of the cross section being examined will not be representative of the preset scratch speed, feed, and so on, but only typical of the braking process alone.

Fundamentals of Scratching Deformation

A large number of parameters interact in a general scratch test (Fig. 4).

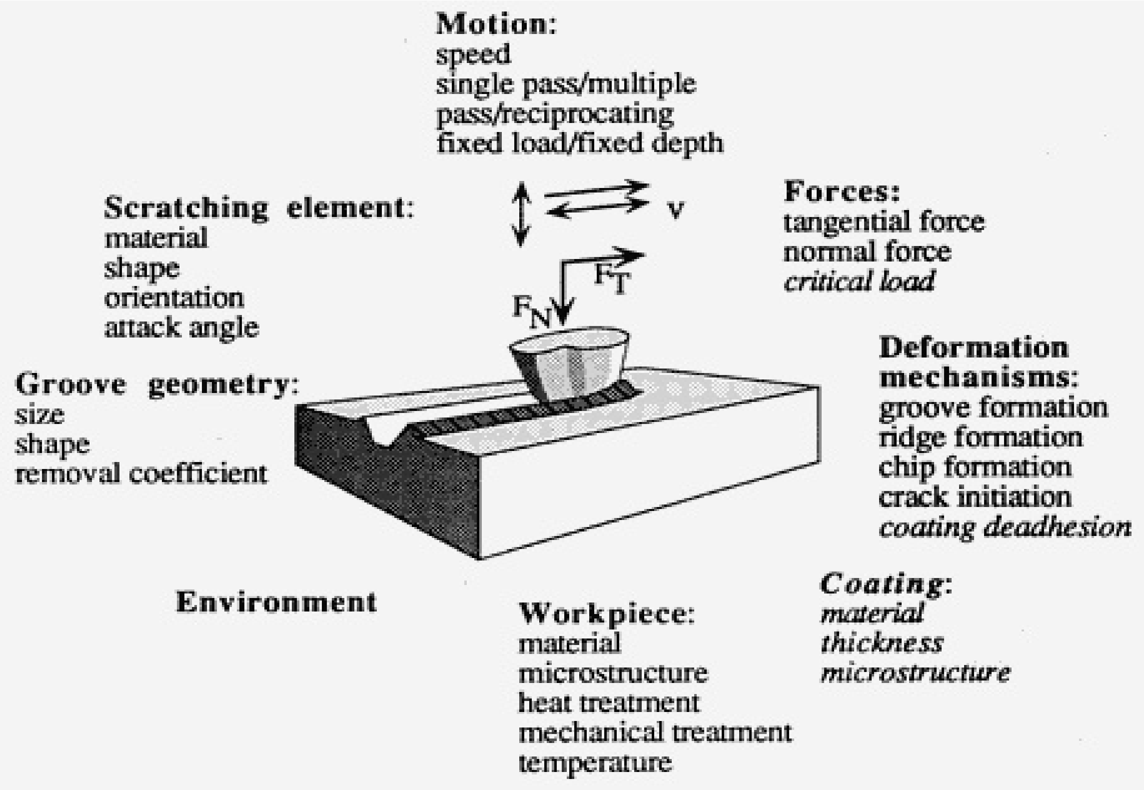


Fig. 4 Schematic showing parameters encountered in scratch testing. F_T, tangential friction force; F_N, normal load, *v*, scratching speed. Parameters in italics are directly associated with surface-coated materials.

The contact geometry at the tip of the grooving element is a function of the following area parameters (Fig. 5, 6):

Area parameter	
Symbol	**Component**
Cross-sectional area	
A_D	Displaced material
A_P	Resulting groove
A_R	Formed ridges
A_W	Removed material
Load-bearing area	
A_{LB}	Vertical projection of the contact area

The first four parameters above refer to cross sections that are perpendicular to the grooving direction (Fig. 5), whereas the fifth parameter is the vertical projection of the contact area (Fig. 6).

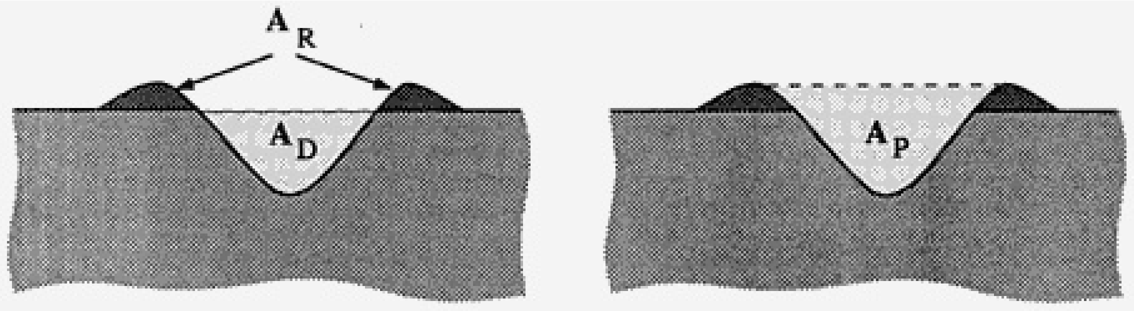


Fig. 5 Cross-sectional area perpendicular to a groove on an initially flat surface. The cross-sectional area A_W of the worn-off material equals A_D - A_R. A_W, cross-sectional area of removed material; A_D, cross-sectional area of displaced material; A_R, cross-sectional area of formed ridges; A_P, cross-sectional area of the resulting groove

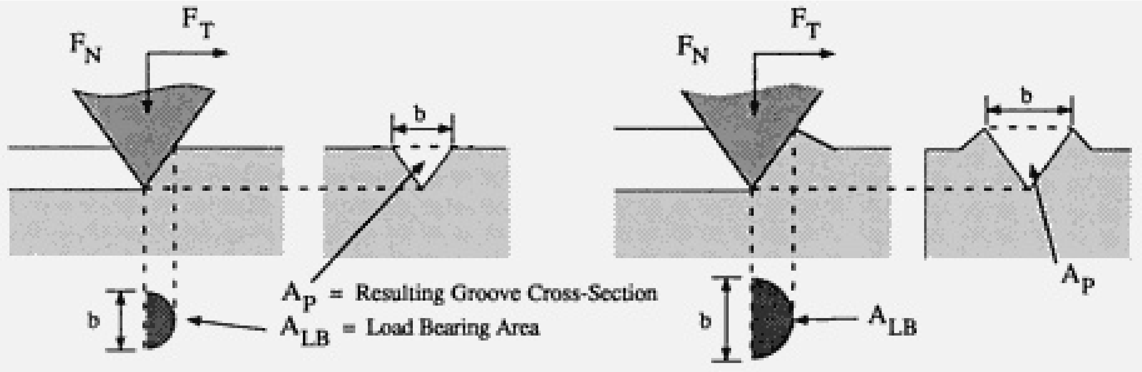


Fig. 6 Effect on load-bearing area, A_{LB}, of scratching with a conical tool. (a) Pure cutting (no ridge formation). (b) A more realistic situation of ridge formation (that is, mixed microplowing and microcutting). b, groove width

Removal Coefficient. The efficiency of material removal of the scratching process can be expressed in terms of the removal coefficient, η, defined as:

$$\eta = \frac{\text{Volume of removed material}}{\text{Volume of displaced material}} = \frac{A_W}{A_D} = \frac{A_D - A_R}{A_D} \quad \textbf{(Eq 1)}$$

Alternative designations for this property are ***degree of wear*** or *abrasive fraction.* The removal coefficient can be used to compare the efficiency of different scratch element shapes or the gage the response of different materials to scratch deformation by forming ridges rather than being worn by chip formation.

Specific Grooving Energy. Another aspect of the efficiency of the material removal process in scratching is the study of the energy required to remove the unit mass of material. This term is often denoted the specific grooving energy, ***e,*** and is defined as:

$$e = \frac{\text{Energy consumed to form the groove}}{\text{Mass of the removed material}} \quad \textbf{(Eq 2)}$$

(Alternatively, the specific grooving energy could be defined as the energy required to remove the unit volume of material.) The specific grooving energy varies with the material as well as with the groove size. Similar to the removal coefficient, the specific grooving energy can be used to relate the performance of different shapes of scratching elements, and to evaluate the response of different materials. Specific grooving energy can also be used to study the influence of groove size on abrasion resistance.

The specific grooving energy is, in fact, the inverse of grindability, a quantity that is often used in the science and practice of grinding. Furthermore, e essentially corresponds to the integral of the specific cutting force k_s, a notation mainly used by the metal cutting industry

Relation of Friction to Scratching Deformation. The process of plastic deformation of material during a scratch test is more complex than that during indentation and will be discussed in this section in a slightly simplified manner. The friction force can be divided into an adhesive and a plowing component (Ref 12, 13). During scratching, the presence of deformation forces that act both parallel and perpendicular to the interface between the stylus and the material being scratched must be determined (Fig. 6).

Under the action of parallel forces, adhesive junctions are sheared and material will flow relative to the stylus. These shear forces constitute the adhesive friction component, F_A, of the total tangential friction force, F_T. Generally, these forces do not necessarily act only in a tangential plane but can also have normal components. In order to simplify the treatment, however, only the adhesive components parallel to the interface will be considered.

The forces acting in a direction normal to the contact surface correspond to the flow pressure of the deformed material as it undergoes deformation. In the present simplified treatment, the vertical flow pressure component is considered to represent the normal load, F_N, and the tangential component constitutes the plowing component of friction, F_P. Provided that the two friction components do not interact, the friction force can be defined as:

$$F_T = F_A + F_P \quad \textbf{(Eq 3)}$$

The sliding contact area can be projected on planes perpendicular to the normal force and on planes perpendicular to the tangential force to give the load-bearing area, A_{LB}, and the scratch cross-sectional area, A_P, respectively (Fig. 6). Those elements of the stylus/workpiece contact area (that correspond to contact against the ridges and against the prow, respectively), are normally included in the projected areas, because they are supposed to contribute to the resistance to deformation. Two stresses may then be defined:

- Scratch hardness, H_S, which is the load per unit load-bearing area (that is, the mean pressure resisting deeper penetration):

$$H_S = \frac{F_N}{A_{LB}}$$ **(Eq 4)**

- Plowing stress, H_P, which is the mean pressure opposing the sliding movement of the stylus due to the plowing process only:

$$H_P = \frac{F_P}{A_P}$$ **(Eq 5)**

-

Under the assumption that the adhesion and plowing can be separated, the friction coefficient, μ, can be written as:

$$\mu = \frac{F_T}{F_N} = \frac{F_A + F_P}{F_N} = \mu_A + \frac{H_P \cdot A_P}{H_S \cdot A_{LB}}$$ **(Eq 6)**

where μ_A is the adhesional component of the friction coefficient. In friction experiments in which the plowing term of the friction is considered to be negligible, Amonton's law is obeyed (that is, the friction coefficient is independent of load). Therefore, it is reasonable to assume that μ_A is generally independent of load and groove size, and that any variation in μ will be due to variations in the plowing term.

In a non-working-hardening material, the plowing stress should be approximately equal to the scratch hardness. However, it has been found (Ref 7) that for annealed polished surfaces the H_P/H_S ratio is ~2 for both microscopic and macroscopic experiments. Furthermore, the H_P/H_S ratio is independent of the scratch width. (It appears that $H_P > H_S$ indicates a higher degree of work hardening in the zone immediately in front of the slider than in the zone underneath it. However, abraded or ground materials are known to be fully work hardened throughout the superficial layer, and thus the work hardening should be the same under and in front of the slider, which would correspond to an H_P/H_S ratio close to unity.)

Assuming that the H_P/H_S ratio has a constant value, *c*, independent of the scratch geometry, the expression for $\mu$is reduced to:

$$\mu = \mu_A + c\ \frac{A_P}{A_{LB}}$$ **(Eq 7)**

The complex load, material, and geometry dependence of this sum has been analyzed for a number of well-defined scratching element shapes of Goddard and Wilman (Ref 13). The friction coefficient was calculated as a function of scratching element shape, orientation, and penetration depth for spherical elements, cones, triangular pyramids, and square pyramids. In addition, Goddard and Wilman also discussed the coupling to actual abrasion.

Scratch Testing of Monolithic Solids

Although scratch testing of bulk material can be performed with numerous objectives, these objectives can be grouped into three categories:

- Scratch hardness evaluation
- Scratching mechanism studies
- Abrasion resistance measurements

Scratch Hardness. Mineralogists and lapidaries have used scratch tests for some time to assess the hardness of stones and minerals.

Mohs Hardness Scale. In 1824, Mohs proposed his famous scale of ten minerals, chosen so that each mineral will scratch all the minerals positioned below it on the scale, but none of the minerals above it on the scale (Ref 14):

Mineral	Hardness	
	Mohs	**HK**
Talc	1	. . .
Gypsum	2	32
Calcite	3	135
Fluorite	4	163
Apatite	5	360[(a)], 430[(b)]
Orthoclase	6	560
Quartz	7	710[(a)], 790[(b)]
Topaz	8	1250
Corundum	9	. . .

(a) Parallel to axis.

(b) Perpendicular to axis

The Mohs hardness of a workpiece or raw material is determined by attempting to scratch it with pieces of the different reference materials. The Mohs hardness number is determined to be the number of the hardest of the reference minerals that fails to produce a scratch on the specimen.

The scale might seem rather arbitrary, but it has been shown by Tabor (Ref 15) to have scientific relevance. It has been shown that in order for one piece of material (shaped as a sharp point) to scratch a flat of another material, it must have an indentation hardness of [ges]1.2 times that of the material of the flat. Thus, any comparative scratch hardness scale such as the Mohs scale must be logarithmic and expressed in indentation hardness numbers with the smallest possible factor between consecutive numbers being 1.2 (that is, every basic material has to be at least 1.2 times harder than the preceding one).

Tabor has shown that the Mohs scratch hardness scale, with the exception of the corundum-to-diamond interval, very closely follows the logarithmic relation to indentation hardness. However, the ratio of indentation hardness for each Mohs number interval is ~1.6. Thus, each increment in the Mohs scale corresponds to a 60% increase in indentation hardness (Knoop or Vickers) and the scale is less accurate than it could ideally be.

State-of-the-Art Scratch Hardness Technology. Scratch hardness was defined above as the load per unit load-bearing area during scratching, taking into account the formation of ridges and a prow (Fig. 6). Because it is generally difficult to measure the load-bearing area during the scratching experiment, it is calculated from the width of the scratch obtained after the test is completed. It is obvious that measurements of the prow that forms in front of the scratching tip poses a problem. As an approximation, it is normally assumed that the prow and the ridges have the same height.

Using this assumption, the scratch hardness can be expressed in terms of the groove width, *b*, for any well-defined scratching tip configuration. This expression varies with tip geometry. For tips of circular cross-sectional area (cones, spheres, and parabolas) the equation is:

$$H_S = \frac{8F_N}{\pi b^2} \quad \textbf{(Eq 8)}$$

For square-base pyramids of leading-edge or leading-face orientation, the equation is:

$$H_S = 4 \frac{F_N}{b^2} \quad \textbf{(Eq 9)}$$

Similar to ordinary quasistatic indentation hardness, scratch hardness is dependent on the shape of the scratching element. This dependence is primarily due to the effect of scratching element shape on prow formation (that is, the size and the extent of upward flow and flow surrounding the tip). Square-base pyramids of leading-face orientation constitute a specific problem because it is not known which faces carry the load: the front face alone, or the front face in combination with the two side faces. Experimental evidence has been presented for both contact situations (Ref 16, 17), and factors such as material selection and experimental design are of major importance. Therefore, this orientation is not recommended for scratch hardness measurements.

One major advantage of scratch hardness testing is that, for very shallow deformations, it is substantially easier to accurately measure the width of a long scratch than to measure the diagonal of an indentation. Long ridges are simply much easier to discern than the outermost corners of an indentation. Thus, scratch testing yields better conditions for testing small structures and very thin surface layers.

Another advantage of scratch hardness relative to indentation hardness measurements is the possibility of studying hardness variations along the scratch. The hardness of different phases (for example, the hardness depth profile caused by deformation hardening or case hardening, or even hardness differences between different crystallographic orientations) can be determined by making one single scratch (Ref 18). The relative hardness changes are easily monitored using a microscope or film records, that is, micrographs. The absolute values at specific positions can be calculated by measuring the groove width and by applying the appropriate hardness formula.

Furthermore, if the hardness value is used to predict abrasion resistance, scratch tests, because they are akin to the abrasion process in many respects, should correlate better than ordinary indentation measurements (Ref 19).

Scratching Mechanisms. Various kinds of scratch tests have proved to be successful in studies of detailed deformation and removal mechanisms. Because both grinding and abrasive wear are functions of the cumulative action of individual grits, scratch testing is a valuable tool to gain understanding of these processes.

The parameters studied include:

- Interrelations between grinding experiments and scratch testing using Rockwell and Vickers diamonds (Ref 6)
- Side wall stripping as a mechanism of material removal during abrasion (Ref 20)
- Mechanisms of carbide removal in white cast iron, including the influence of multiple passes over a preworn surface (Ref 11)
- Transitions between different chip shapes as the one angles of conical scratch tools are varied
- Formation and effect of a stagnant zone ahead of the scratch tool (Ref 5)

In general, it has been determined that material is more easily removed with repeated traversals as opposed to a single scratch across a polished surface (Ref 6).

Size Effects. In an investigation of scratch deformation *in situ* in a scanning electron microscope, Gane and Skinner (Ref 7) used a spherically rounded tungsten stylus with a tip radius of ~1 μm (~40 μin.) to carry out microfriction experiments on gold and copper materials. They determined that H_S rises steeply with decreasing groove sizes. For a groove width of ~0.3 μm (~12 μin.), hardness values were three to five times the bulk hardness. This size effect is also well known from static indentation investigations, but was found to be much more pronounced under scratch deformation testing.

The H_P/H_S ratio was ~2 and independent of the scratch width. At the smallest scratch sizes, the plowing stress H_P was even found to approach the theoretical strength of the perfect crystal. It would seem likely, therefore, that a limit has been reached, and that any further reduction in size would not lead to any appreciable increase in hardness.

The coefficient of friction was found only to exhibit the size dependence generated by the term A_P/A_{LB}, as predicted by Eq 7. This is a direct result of the proportional increase in H_P and H_S for small groove sizes.

Abrasion Resistance. Scratch testing has been adopted to investigate or rank the resistance to abrasive and grooving wear. Two different approaches are described here:

- One method employs scratch testing to estimate the removal coefficient of materials and uses this number and the hardness of the material in the fully strain hardened condition to make a criterion better correlated to the abrasion resistance than a plain hardness value
- Second method employs the concept of specific grooving energy to obtain rankings of competitive grooving resistant materials

Correlation Between Scratch Hardness and Abrasion Resistance. It is known that the abrasion resistance of pure metals is very closely proportional to their indentation hardness, *H*. However, this in not necessarily the case for alloys in which the hardness values are varied by heat treatments (Fig. 7). However, Zum Gahr (Ref 21, 22) has shown that the proportionality is more prominent if an increase in hardness is obtained by heavy work hardening of the material (H_{def} in Fig. 7b), instead of by heat treatments. It is possible to assess H_{def}, for example, by measuring the hardness of microchips formed under abrasion. Additional improvement of the linear relation is achieved by using the H_{def} value and further compensating for the material removal (that is, considering the ability of the material to respond to the engagement of a scratching tool by forming ridges rather than wear fragments or microchips). Because it is difficult to measure the removal coefficient during abrasion, Zum Gahr employed a scratch test to make an estimate. Single scratches on initially flat surfaces were made using a scratch diamond with an attack angle of 90°, a tip radius of 8 μm (320 μin.) and loaded by 2 N (0.2 kgf). The cross-sectional geometry of the grooves was subsequently measured and the removal coefficient calculated according to Eq 1. The acquired value, f_{ab}, is not identical to the average removal coefficient of real abrasion, which corresponds to more complicated abrasive element shapes and is performed on surfaces of irregular topography, and so on. Furthermore, because the f_{ab} value is determined by single-pass scratching, the ability of the ridges to withstand deformation from a subsequent scratching event is not considered. Still, this estimate of the removal coefficient has been shown to have good relevance, and a substantially improved proportionality is shown by plotting the abrasion resistance versus H_{def}/f_{ab} (Fig. 7c). Comparing Fig. 7(b) and 7(c) also shows that if two materials have the same H_{def}, the composition with the lowest f_{ab} value will have the best abrasion resistance.

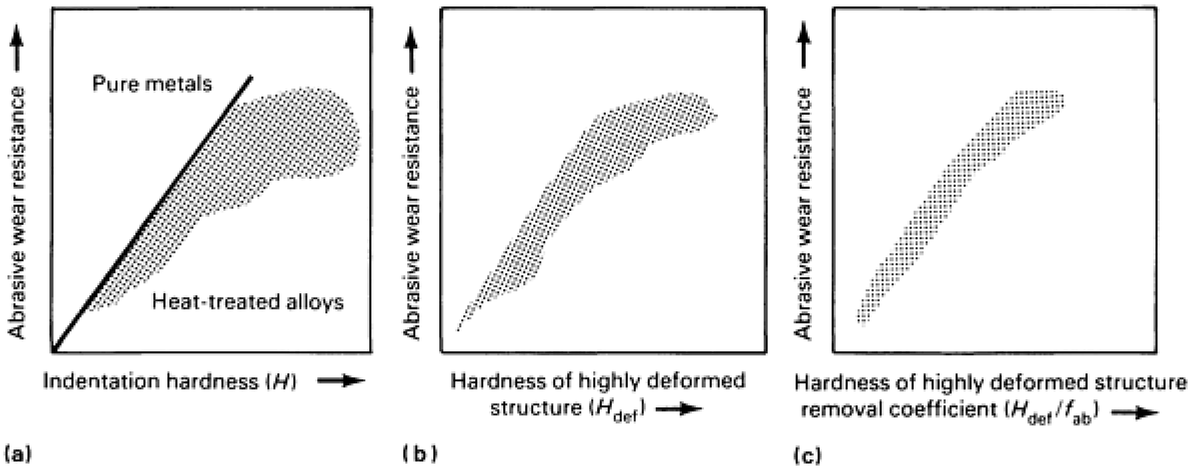


Fig. 7 Plots showing the correlation between abrasive wear resistance and selected hardness parameters. (a) Indentation hardness. (b) Hardness of a highly deformed structure. (c) Hardness of a highly deformed structure divided by the removal coefficient, f_{ab}. The individual results are continued to the shaded areas.

Specific Grooving Energy. Vingsbo *et al.* (Ref 1, 2) have used a pendulum tester and the concept of specific grooving energy to rank materials relative to resistance against grooving wear. In the pendulum test, a series of grooves of increasing size are produced, the corresponding mass losses, *w*, are weighed, and the consumed amounts of energy, *E*, are recorded. The specific grooving energy is defined as:

$$e = \frac{E}{w} \quad \textbf{(Eq 10)}$$

Figure 8 is a plot of log *e* versus log *w* and is used to rank different materials, based on the assumption that a high specific energy is directly related to a high resistance against grooving wear.

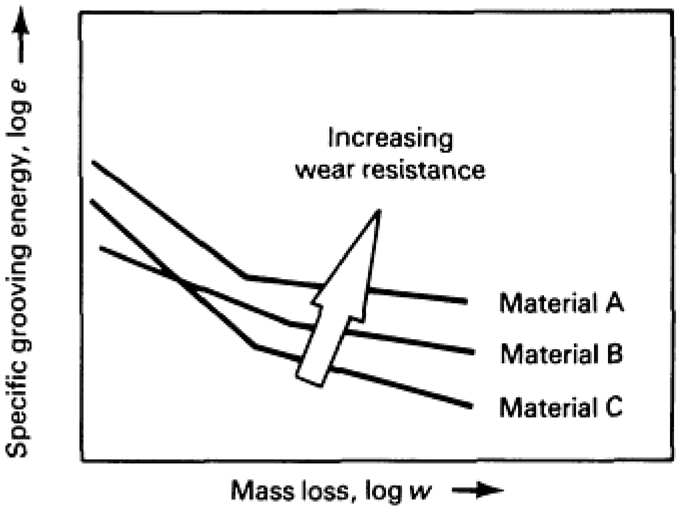


Fig. 8 Plot of specific grooving energy versus mass loss (abrasive wear) for three different materials

Based on Fig. 8, *e* will vary with *w* (groove size) not only relative to material, but also relative to groove size (that is, the character of the abrasive attack). Thus, a material that has the best resistance against coarse grooving wear does not necessarily have the same ranking relative to careful grinding or polishing.

Additional advantages of the pendulum technique are:

- Low experimental scatter
- A wide groove size interval
- Scratching speeds typical of wear parts in many service applications (for example, mining equipment, road grading blades, or excavator teeth)

Scratch Adhesion Testing of Thin Hard Coatings

Thin, hard wear-resistant coatings are today successfully employed in various engineering applications. However, if the adhesion of the coating to the substrate is insufficient, premature failure of the coated part may occur because of coating detachment by interfacial delamination or fracture. Thus, evaluation of coating adhesion is a key issue, and relevant test methods are sorely needed. Despite the publication of more than twenty methods on adhesion testing available in the literature, the authors of three independent review papers (Ref 23, 24, 25) on adhesion testing have all arrived at the conclusion that the only methods applicable to hard well-adherent coatings are the laser photoacoustic shockwave test (Ref 26) and the scratch adhesion test (Ref 27). Today, the scratch test is by far the most widely used because it is easy to perform and also yeilds comparatively fast results. A contributing factor to the popularity of the scratch test is the availability of commercial equipment (Fig. 9) from several manufacturers (Ref 28, 29). As of today, the scratch adhesion test is not standardized but standard initial work in terms of round-robin experiments have been performed (Ref 30, 31).

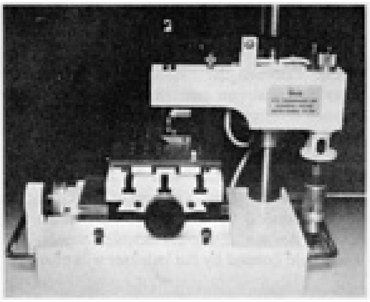


Fig. 9 A typical commercial scratch testing apparatus.

Test Equipment and Procedures. During the scratch test, a diamond stylus is made to slide over the coated surface (Fig. 10). The applied normal force, F_N, is increased stepwise or continuously (Fig. 11) until the coating is detached. The normal force that produces coating failure is known as the "critical normal force," $F_{N,C}$, and is used as a measure of the adhesion. Figures 2(a) and 2(b) show typical resulting scratches for a normal force $<F_{N,C}$ and $>F_{N,C}$, respectively. The onset of peeling can be monitored by optical microscopy, acoustic emission, and friction force measurements (Ref 32), and it has been suggested that the $F_{N,C}$ transition is directly associated with the sudden increase in the friction force and acoustic emission readings (Ref 28, 29).

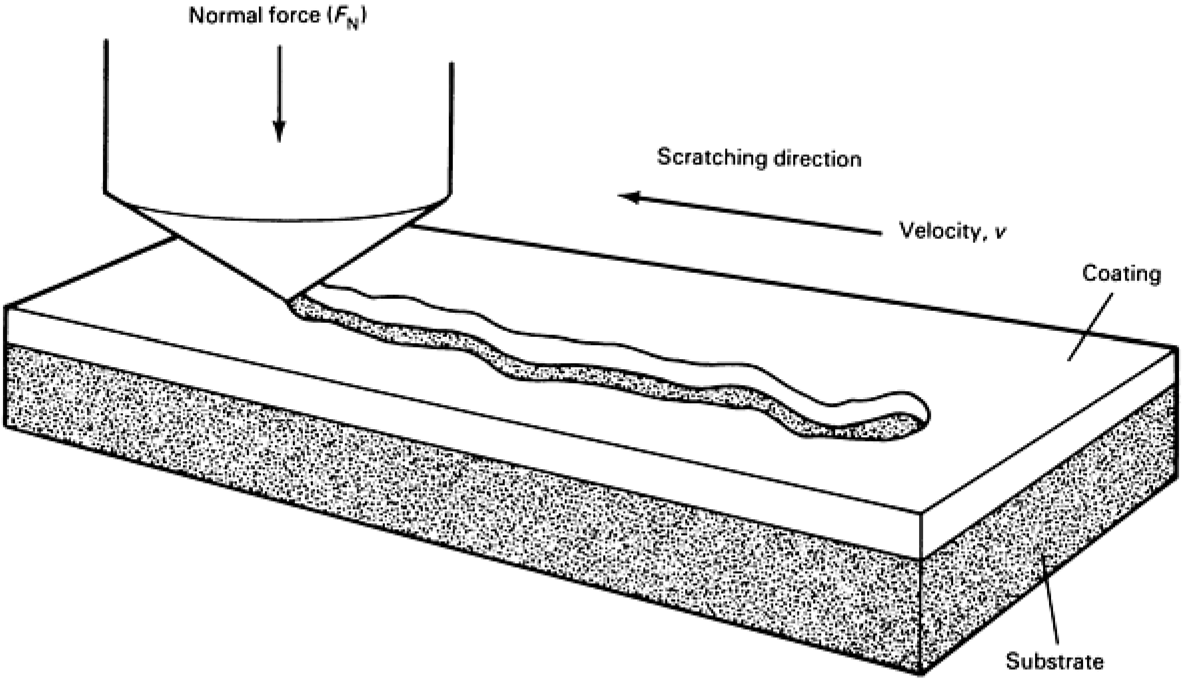


Fig. 10 Schematic showing typical deformation generated by diamond stylus sliding over a coated surface

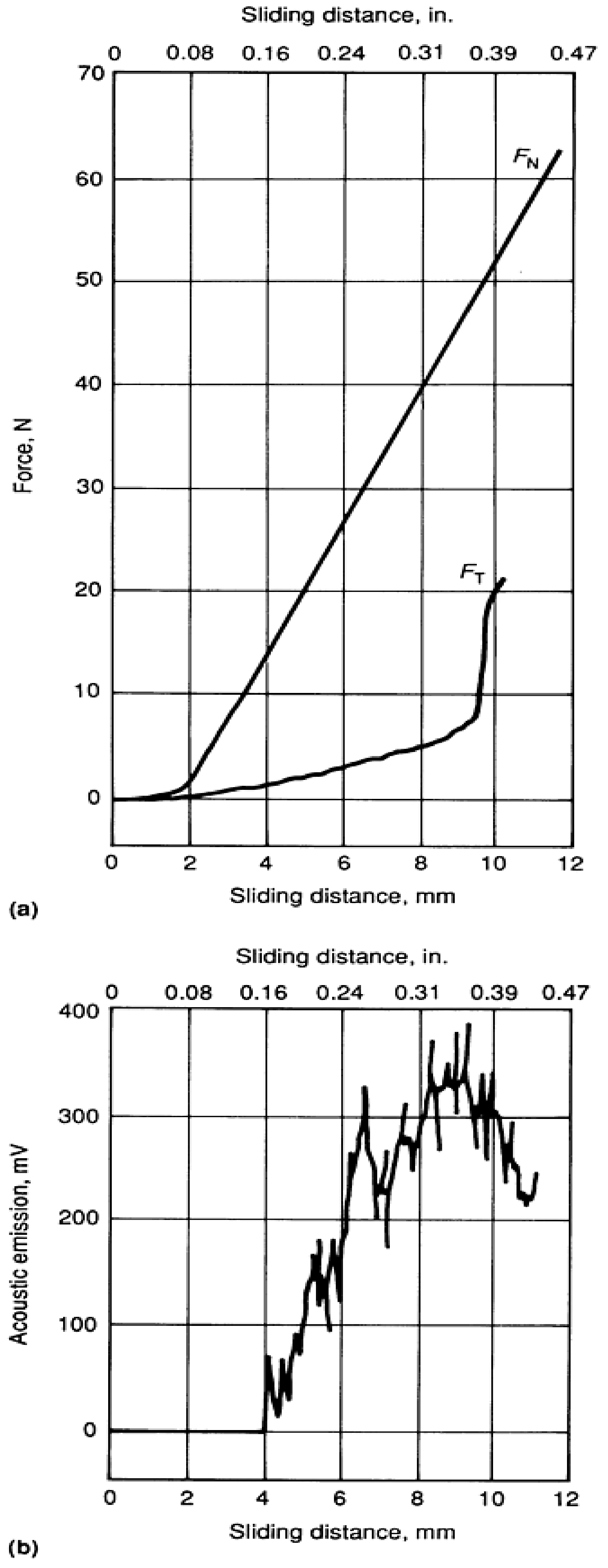


Fig. 11 Monitoring of peeling of a coating during a scratch test by two methods. (a) Friction force measurement. (b) Acoustic emission testing. Source: Ref 23

The most straightforward method for monitoring scratch testing in continuous load tests is optical microscopy. The distance from the start of the scratch to the failure site can be measured and directly related to the normal force. In principle, any optical microscope can be used to perform this measurement, but a microscope integrated with the test equipment is obviously advantageous.

The friction force detector (a strain gage transducer) detects the change in friction force when the coating is detached. Valli *et al.* (Ref 33) have found this method to be more sensitive in detecting coating failure, particularly for very thin (0.1 μm, or 4 μin.) coatings, than the other methods. A typical friction force trace is shown in Fig. 11. The acoustic detector (an accelerometer, basically measuring mechanical vibrations) is mounted above the stylus and detects the onset of coating detachment through the accompanying shock wave bursts (Fig. 11). However, neither of the latter two detection methods has proved completely reliable. Therefore, a posttest examination of the scratch should always be performed (Ref 23, 34).

Attempts have been made to line $F_{N,C}$ with the adhesion of the coating, using approaches ranging from pure plasticity theory (Ref 35) over elastic-plastic models (Ref 36, 37, 38) to models based on the release of strain energy during the removal of a coating (Ref 39, 40). The latest and most successful attempt has been made by Bull *et al.* (Ref 4), using the relation:

$$F_{N,C} = \frac{A_1}{\nu_c \mu_C} \left(\frac{2E_c W_{12}}{t} \right)^{1/2} \quad \textbf{(Eq 11)}$$

between adhesion and $F_{N,C}$. Here A_1 is the cross-sectional area of the scratch; v_c is the Poisson's ratio of the coating material; μ_C is the coefficient of friction at $F_{N,C}$; E_c is the Young's modulus of the coating; t is the thickness of the coating; and W_{12} is the work of adhesion. W_{12} is given by:

$$W_{12} = \gamma_1 + \gamma_2 - \gamma_{12} \quad \textbf{(Eq 12)}$$

where γ_1 and γ_2 are the specific surface energies of coating and substrate, respectively, and γ_{12} is the interfacial specific free energy.

However, the work of Bull *et al.* concentrates on one particular type of failure (ahead of the moving indenter), and the results will only apply when tensile stresses, which are normal to the surface, cause coating detachment. Because a number of possible failure mechanisms have been identified (Ref 34, 37), the number of coating/substrate systems that can be studied using this approach is limited.

Parameters Affecting $F_{N,C}$. Because the $F_{N,C}$ value is affected by numerous parameters related both to the testing conditions and to the coating/substrate systems being investigated (Ref 23, 28, 29, 34, 42), it cannot be directly related to the strength of the coating/substrate interface. In addition, the correlation between scratch test data and actual coating performance in practical applications is often poor. Attempts to quantitatively express adhesion in terms of critical normal force are also made more difficult. However, the influence of some parameters, such as temperature, humidity, and so on, still remains to be investigated.

Substrate Hardness and Coating Thickness. The influence of these parameters on the $F_{N,C}$ value has been thoroughly investigated (Ref 23, 28, 34, 43, 44, 45). It has been shown that the $F_{N,C}$ values increases both with increasing substrate hardness and with increasing coating thickness (Fig. 12).

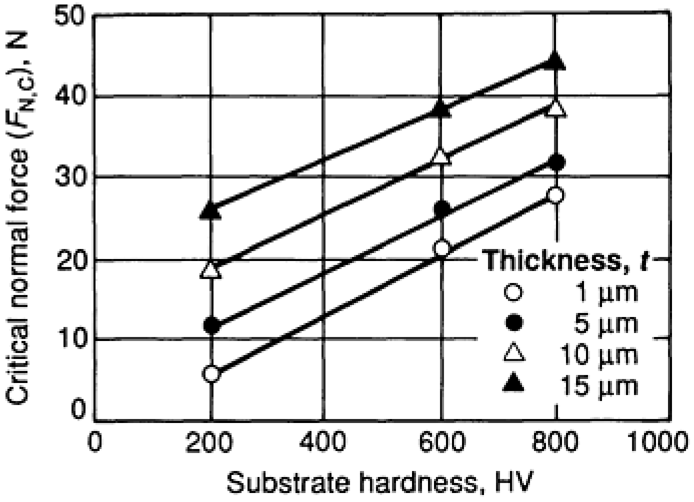


Fig. 12 Plot of critical normal force versus substrate hardness for selected thicknesses of chemical vapor deposition (CVD) TiC coatings on various steels. Source: Ref 28

Coating Surface Roughness. During scratch tests on commercially available inserts for cutting tools, it has been shown that the critical normal face may depend on the coating surface roughness. In particular, $F_{N,C}$ values determined on coated inserts with a high surface roughness (R_a ≥0.3 mm, or 12 μin., where R_a is the arithmetic average surface roughness) have no significance because such coatings have failures that occur mainly within the coating itself (cohesive failures) (Ref 42) and not at the coating/substrate interface.

Substrate Roughness Prior to Coating. The exact influence of substrate roughness on the $F_{N,C}$ value is difficult to determine because differences in surface roughness affect both the efficiency of the cleaning operation used and the adhesive properties of the coatings. However, studies have found (Ref 42) that the critical normal force have decreases with increasing surface roughness of the substrate (Fig. 13).

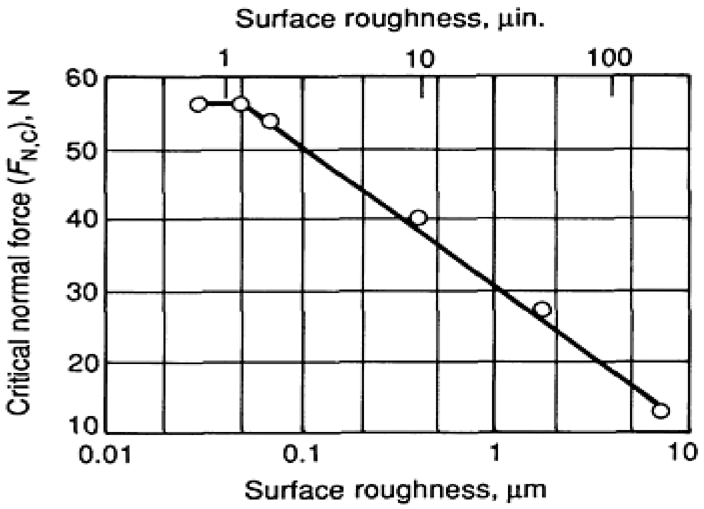


Fig. 13 Plot of critical normal force versus substrate surface roughness obtained for physical vapor deposition (PVD) TiN-coated high-speed steel (HSS) samples that have been subjected to different surface preparations.

Coating thickness was 2 μm (80 μin.). Source: Ref 42

Loading Rate and Scratching Speed. $F_{N,C}$ has been found to be independent of both loading rate (dF_N/dt) and scratching speed (dx/dt), provided that the ratio dF_N/dx remains constant (a "standard" dF_N/dx ratio is 10 N/mm). If the loading rate, dF_N/dx is kept constant, $F_{N,C}$ decreases when the scratching speed increases (Fig. 14a); when the scratching speed remains constant, $F_{N,C}$ increases with increasing loading rate (Fig. 14b). Furthermore, it has also been noted that a high loading rate results in a high experimental scatter.

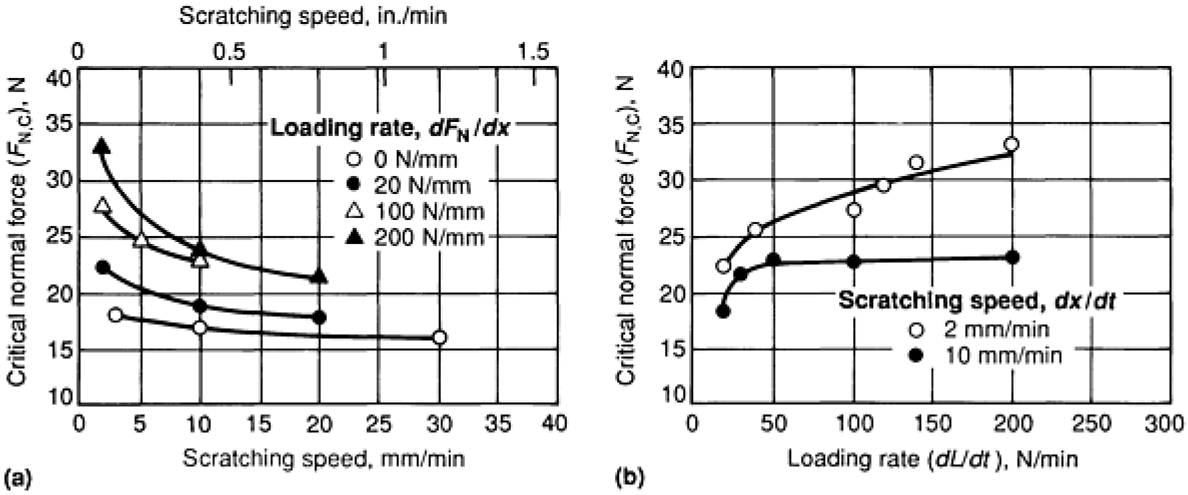


Fig. 14 Variation in $F_{N,C}$ measured on CVD TiC-coated steel. (a) Plot of $F_{N,C}$ versus scratching speed for selected loading rate values. (b) $F_{N,C}$ versus loading rate for selected scratching speeds. Source: Ref 42

Diamond Tip Radius. Theoretically, $F_{N,C}$ should be proportional to R^2 (Ref 42), where R is the diamond tip radius. However, experimental results vary from $R^{0.85}$ to $R^{1.55}$ for different dF_N/dx ratios and coating/substrate system. The most common R value is 200 μm (0.008 in.), which is given by the geometry of the HRC hardness test. Hamersky (Ref 46) has shown this to be the lowest value of R possible for accurate test results. However, his work was limited to aluminum films on glass substrates, and some doubt still exist about its relevance for hard coatings on hard substrates.

Friction Between Indenter and Coating. The friction between the indenter and the coating plays an important role (Ref 23, 28, 33, 42) because different friction coefficients yield different stress fields in the specimen ad thus will have different effects on the critical normal force.

Diamond Tip Wear. Obviously, the condition of the diamond tip will affect the scratch test results. As the diamond indenter wears, the stress field created by the indenter will change.

References

1. O. Vingsbo and S. Hogmark, Single-Pass Pendulum Grooving--A Technique for Abrasive Testing, *Wear,* Vol 100, 1984, p 489-502
2. U. Bryggman, S. Hogmark, and O. Vingsbo, Mechanisms of Gouging Abrasive Wear of Steel Investigated with the Aid of Pendulum Single-Pass Grooving, *Wear,* Vol 112, 1986, p 145-162
3. T.C. Buttery and M.S. Hamed, Some Factors Affecting the Efficiency of Individual Grits in Simulated Grinding Experiments, *Wear,* Vol 44, 1977, p 231-241
4. D. Graham and R.M. Bual, An Investigation into the Mode of Metal Removal in the Grinding Process, *Wear,* Vol 19, 1972, p 301-314
5. Y. Kita and M. Ido, The Mechanism of Metal Removal on Abrasive Tool, *Wear,* Vol 47, 1978, p 185-193
6. T.C. Buttery and J.F. Archard, Grinding and Abrasive Wear, *Proc. Inst. Mech. Eng.* (1970-1971), 185 27/71

7. N. Gane and U. Skinner, The Friction and Scratch Deformation of Metals on a Micro Scale, *Wear,* Vol 24, 1973, p 207-217
8. L. Åhman and Å. Öberg, Mechanisms of Micro-Abrasion--*In situ* Studies in SEM, *Proceedings of the International Conference on Wear of Materials,* 1983, p 112-120
9. K. Kato, K. Hokkirigawa, T. Kayaba, and Y. Endo, Three Dimensional Shape Effect on Abrasive Wear, *J. Tribol. (Trans. ASME),* Vol 108, 1986, p 346-351
10. S.J. Calabrese, F.F. Ling, and S.F. Murray. Dynamic Wear Tests in the SEM, *ASLE Trans.,* Vol 26 (No. 4), 1982, p 455-465
11. S.V. Prasad and T.H. Kosel, A Study of Carbide Removal Mechanisms during Quartz Abrasion. I: In situ Scratch Test Studies, *Wear,* Vol 92 (No. 2), 1983, p 253-268
12. F.P. Bowden and D. Tabor, *The Friction and Lubrication of Solids,* Oxford University Press, 1950
13. J. Goddard and H. Wilman, A Theory of Friction and Wear during the Abrasion of Metals, *Wear,* Vol 5, 1962, p 114-135
14. E. Rabinowicz, *Friction and Wear of Materials,* John Wiley & Sons, 1966
15. D. Tabor, The Hardness of Solids, *Rev. Phys. Technol.,* Vol 1, 1970, p 145-179
16. T. Hisakado, On the Mechanisms of Contact between Solid Surfaces, *Bull. JSME,* Vol 13 (No. 55), 1970, p 129-139
17. A. Broese van Groenou, N. Maan, and J.D.B. Veldkamp, Scratching Experiments on Various Ceramic Materials, *Philips Res.,* Vol 30, 1975, p 320-359
18. H.D. Beurs, G. Minholts, and J.T.M.D. Hosson, Scratch Hardness and Wear Performance of Laser Melted Steels: Effects of Anisotropy, *Wear,* Vol 32, 1989, p 59-75
19. P.J. Blau, Relationships between Knoop and Scratch Micro-Indentation Hardness and Implications for Abrasive Wear, *Microstr. Sci.,* Vol 12, 1985, p 293-313
20. A.A. Torrance, A New Approach to the Mechanics of Abrasion, *Wear,* Vol 67 (No. 2), 1981, p 233-257
21. K.H. Zum Gahr, *Microstructure and Wear of Materials,* Tribology Series, Elsevier, Amsterdam, 1987
22. K.H. Zum Gahr, Modelling of Body Abrasive Wear, *Wear,* Vol 124, 1988, p 87-102
23. J. Valli, *J. Vac. Sci. Technol. A,* Vol 4 1986, p 3007
24. A.J. Perry, *Surf. Eng.,* Vol 2, 1986, p 183
25. D.S. Rickerby, *Surf. Coat. Technol.,* Vol 36, 1988, p 541
26. J.L. Vossen, *Adhesion Measurement of Thin Films, Thick Films and Bulk Coatings,* STP 640, K.L. Mittal, Ed., ASTM, 1987, p 122
27. A.J. Perry, *Thin Solid Films,* Vol 107, 1983, p 167
28. P.A. Steinmann and H.E. Hintermann, *J. Vac. Sci. Technol. A,* Vol 3, 1985, p 2394
29. J. Valli, and Mäkelä, *Wear,* Vol 115, 1987, p 215
30. A.J. Perry, J. Valli, and P.A. Steinmann, *Surf. Coat. Technol.,* Vol 36, 1988, p 559
31. H. Ronkainen, S. Varjus, K Holmbreg, K.S. Fancey, A.R. Pace, A. Matthews, B. Matthews, and E. Broszeit, Paper presented at the 16th Leeds-Lyon Symposium on Tribology, 5-8 Sept 1989 (Lyon)
32. J. Sekler, P.A. Steinmann, and H.E. Hintermann, *Surf. Coat. Technol.,* Vol 36, 1988, p 519
33. J. Valli, U. Mäkelä, A. Matthews, and W. Murawa, *J. Vac. Sci. Technol. A,* Vol 3, 1985, p 2411
34. P. Hedenqvist, M. Olsson, S. Jacobson, and S. Söderberg, *Surf. Coat. Technol.,* Vol 41, 1990, p 31
35. P. Benjamin and C. Weaver, *Proc. R. Soc. (London) A,* Vol 254, 1960, p 163
36. C. Weaver, *J. Vac. Sci. Technol.,* Vol 12, 1975, p 18
37. P.J. Burnett and D.S. Rickerby, *Thin Solid Films,* Vol 154, 1987, p 403
38. P.J. Burnett and D.S. Rickerby, *Thin Solid Films,* Vol 157, 1988, p 233
39. M. Laugier, *Thin Solid Films,* Vol 76, 1981, p 289
40. M. Laugier, *Thin Solid Films,* Vol 117, 1984, p 243
41. S.J. Bull, D.S. Rickerby, A. Matthews, A. Leyland, A.R. Pace, and J. Valli, *Surf. Coat. Technol.,* Vol 36,

1988, p 503

42. P.A. Steinmann, Y. Tardy, and H.E. Hintermann, *Thin Solid Films,* Vol 154, 1987, p 333
43. P. Laeng and P.A. Steinmann, *Proceedings of the 8th International Conference on Chemical Vapor Deposition,* 15-18 Sept 1981 (Paris), Electrochemical Society, 1981, p 723
44. B. Hammer, A.J. Perry, P. Laeng, and P.A. Steinmann, *Thin Solid Films,* Vol 96, 1982, p 45
45. E. Hummer and A.J. Perry, *Thin Solid Films,* Vol 101, 1983, p 243
46. J. Hamersky, *Thin Solid Films,* Vol 3, 1969, p 263

Surface Temperature Measurement

Francis E. Kennedy, Jr., Thayer School of Engineering, Dartmouth College

Introduction

SLIDING FRICTION results in a loss of mechanical energy, and past studies have shown that the vast majority of frictional energy is transformed into heat. This frictional energy dissipation takes place in the immediate vicinity of the real area of contact, where frictional interactions occur. The transformation of frictional energy into heat, a process known as "frictional heating," is responsible for increases in the temperatures of the sliding bodies, especially in the contact region. Frictional heating and the resultant contacting temperatures can significantly influence the tribological behavior and failure of sliding components. Surface and near-surface temperatures can become high enough to cause changes in the structures and properties of the sliding materials, oxidation of the surface, and possibly even melting of the contacting solids. Therefore, it is important to be able to predict and/or measure the temperatures of sliding contacts. Analytical methods for surface temperature determination are discussed in the article "Frictional Heating Calculations" in this Volume. This article will concentrate on experimental techniques for the measurement of surface and near-surface temperatures in contacting bodies.

A description of the geometric and temporal conditions under which such temperatures occur is useful background information. As shown in Fig. 1, there are three levels of temperature in sliding contacts. The highest contact temperatures, T_c, occur at the small (perhaps on the order of 10 μm diam) contact points between surface roughness peaks or asperities on the sliding surfaces. These temperatures can be above 1000 °C (1830 °F), but last only as long as the two asperities are in contact, possibly less than 10 μs. The asperity contacts are often confined to a small portion of the surface of the bodies, termed the "contact patch." An example of this is a typical elliptical Hertzian contact area several hundred micrometers in length between the ball and the ring in a rolling-element bearing. At any instant, there are usually several short-duration contact temperatures (T_c) at the various asperity contact points within a contact patch. The integrated (in space and time) average of the temperatures of all points within the contact area is the "mean contact temperature" (T_m). The mean contact temperature can be above 500 °C (930 °F) for severe sliding cases, such as in brakes, but is usually much lower. The temperature diminishes as the distance from the contact patch increases, and it generally reaches a rather modest "bulk volumetric temperature" (T_b) several millimeters into the contact bodies. That temperature is generally less than 100 °C (210 °F). It might be noted that the flash temperature originally discussed by Blok (Ref 1) and used in the article "Frictional Heating Calculations" in this Volume is defined as $T_f = T_c - T_b$.

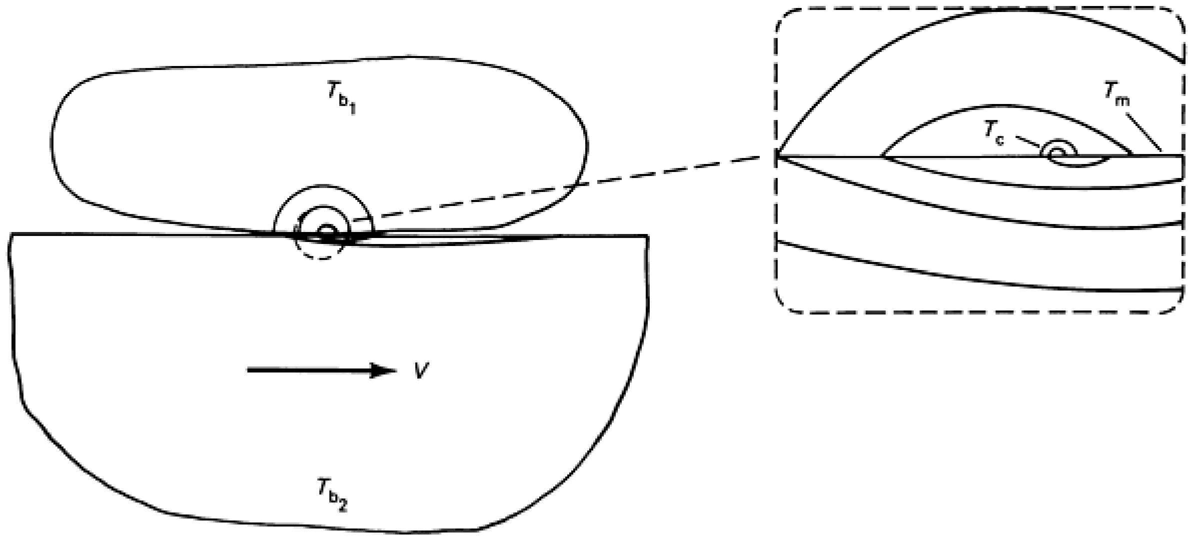


Fig. 1 Schematic diagram of temperature distribution (isotherms) around sliding contacts

Many experimental techniques for surface temperature measurement have been used, with varying degrees of success, to study the temperatures that result from frictional heating. Most of the techniques have geometric or temporal limitations that prevent their use for measuring all three levels of sliding surface temperature. The more successful techniques will be described below.

Metallographic Techniques

Examination of the microstructures of sections of bodies that have undergone frictional heating can provide information about the temperatures the bodies witnessed in service. Such metallographic techniques generally measure the microstructural changes that result from the surface and near-surface temperatures. These changes can be detected after metallurgical sectioning of the sliding body in a plane normal to the sliding direction. For some materials, etching of the near-surface region of the cross-sectioned body reveals a visible change in microstructure (Ref 2). For other materials, microhardness surveys are effective for determining the near-surface temperature distribution that the material witnessed in service (Ref 3). In either case, the temperature contours are constructed by comparing the hardness or structural appearance variations with those of standard reference specimens heat treated to known temperatures for known lengths of time (Ref 2).

An example of the use of this microstructural technique is shown in Fig. 2 (Ref 3). Figure 2 shows an etched cross section of a high-speed steel cutting tool after cutting iron for 30 s at a speed of 3 m/s (10 ft/s). The material beneath the cutting face of the tool was tempered as a result of frictional heating. Figure 3 illustrates the temperature contours developed by comparing the tempering effects in the tool with those in specimens of the same material (M34 high-speed steel) that were heated to known temperatures for the same length of time.

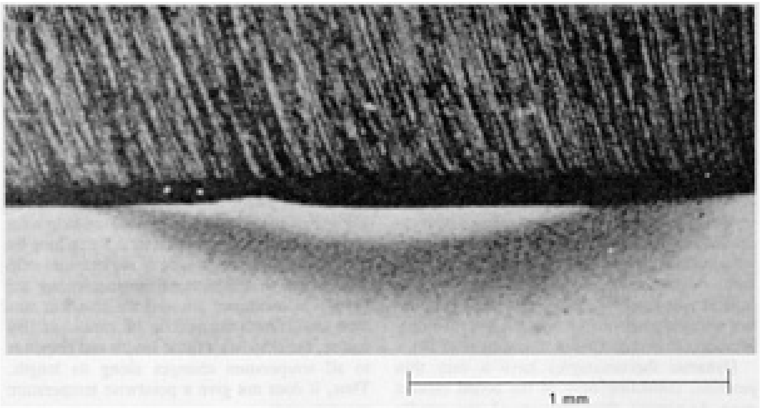


Fig. 2 Metallographic section through the cutting edge of a high-speed steel tool used for cutting iron for 30 s, with adhering chip. Section was etched in 2% nital for 30 s and shows heated region beneath rake face. Cutting speed: 3 m/s (10 ft/s). Source: Ref 3

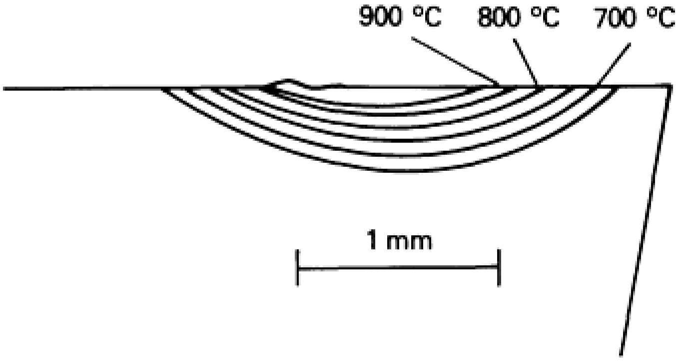


Fig. 3 Temperature contours for the cutting tool shown in Fig. 2. Source: Ref 3

Metallographic techniques generally require destruction of the sliding body for sectioning. Such postmortem investigations can yield substantial information about the mean surface and bulk volumetric temperatures that occurred in the sliding body during service. They can be used successfully only with materials that undergo a known change in microstructure or microhardness at the temperatures encountered in sliding. For the temperature contours to be accurate, the microstructural transformation of the material must be relatively unaffected by the contact stresses and temperature transients that occur at the sliding interface. Metallographic techniques cannot measure short-duration contact temperatures (T_c), nor can they measure temperature at a given instant during the sliding process.

Thermocouples and Thermistors

Thermocouples are probably the most commonly used sensors for measuring the temperatures caused by frictional heating. Their operation is based on the findings of Seebeck, who in 1821 demonstrated that a specific electromotive force (emf) potential exists as a property intrinsic to the composition of a wire whose ends are kept at two different temperatures. The simplest measuring circuitry for thermocouple thermometry involves wires of two dissimilar metals connected so as to give rise to a total relative Seebeck potential. This emf is a function of the composition of each wire and the temperatures at each of the two junctions. This circuit can be well characterized such that, if one junction is held at a known reference temperature, the temperature of the other "measuring" junction can be inferred by comparing the measured total emf with an empirically derived calibration table (Ref 4).

Embedded Subsurface Thermocouples. Generally, when the thermocouple technique is to be used to measure contact temperature, a small hole is drilled into a noncontacting surface of the stationary component of a frictional pair. The hole may extend to, or just beneath, the sliding surface. Cement is put in the hole; ceramic cement is used for components that encounter high sliding temperatures, and polymeric adhesive, such as epoxy, is used for lower-

temperature components. A small thermocouple is then inserted in the hole so that its measuring junction rests either at or just beneath the sliding surface. It is held in position by the cement, which also serves to electrically insulate the thermocouple wires from the surrounding material. Figure 4 is a diagram of a typical embedded thermocouple installation (Ref 5). Several such thermocouples can be embedded at different depths and a various locations along the sliding path to obtain information about surface temperature distribution and temperature gradients (Ref 6). Monitoring the thermocouples throughout a sliding interaction allows changes in surface temperature to be deduced.

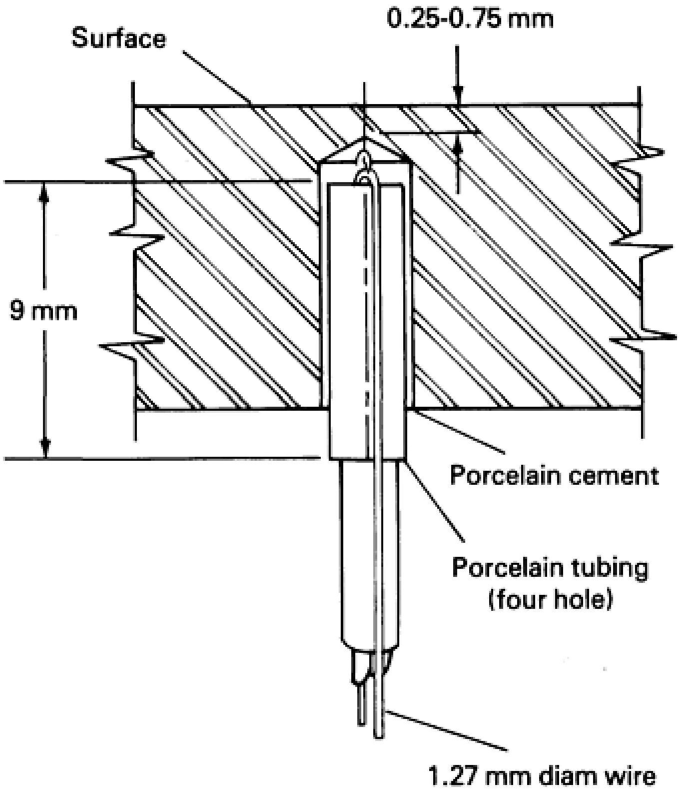


Fig. 4 Diagram of a thermocouple embedded in a brake pad. Source: Ref 5

Embedded thermocouples have been found to provide a good indication of the transient changes in frictional heat generation that accompany contact area changes (Ref 5, 6, 7). They cannot, however, provide a true indication of surface temperature peaks. Because of their mass and distance from the points of intimate contact where heat is being generated, subsurface thermocouples have a limited ability to respond to flash temperatures. A thermocouple can be made part of the sliding surface by placing it in a hole that extends to the surface and then grinding the thermocouple even with the surface. Even in that case, however, the finite mass of the thermocouple junction prevents it from the responding to flash temperatures of very short duration (Ref 7). Although these problems are not as severe with fast-response microthermocouples, the best use for embedded thermocouples is for measuring bulk temperatures within the sliding bodies. These bulk temperatures can be used effectively in determining boundary conditions for an analytical study or for calculating the distribution of frictional heat between the two contacting bodies (Ref 8). This is accomplished most easily if temperatures are measured at several depths beneath the sliding surface, enabling the determination of heat flux values.

Contact thermocouples consist of two separate insulated wires embedded in one of the sliding components but exposed at the contact surface. The deformation and frictional heating at the contact zone join the wires together, creating a thermocouple junction that responds to contact surface temperature. Such thermocouples have been used in several experimental studies of grinding temperature in the former Soviet Union (Ref 9) and elsewhere (Ref 10). An example of such a thermocouple is shown in Fig. 5 (Ref 10). In this design, separate flattened thermocouple wires are separated by insulating sheets of mica and are sandwiched between two halves of a split specimen (the workpiece). It is essential that the two thermocouple electrodes be insulated from each other and from the workpiece so that the thermal emf is not short-circuited (Ref 10). When sliding occurs at the hot junction, the two electrodes are welded together by the heat generated in the contact zone. The resulting thermocouple then indicates the temperature of the hot junction at the sliding interface.

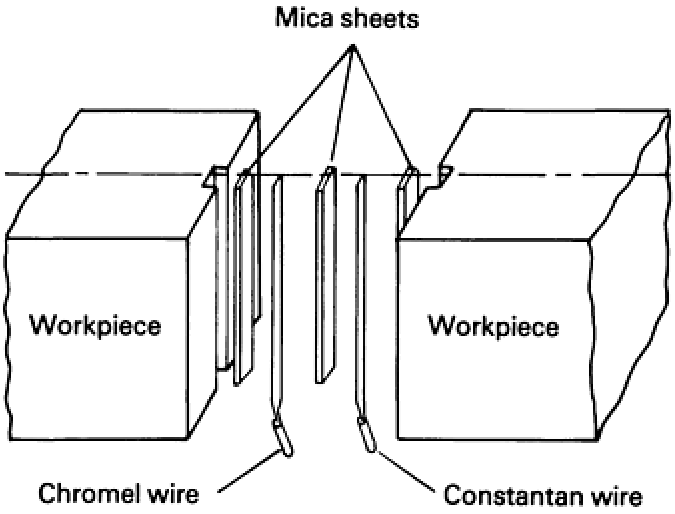


Fig. 5 Schematic diagram of a sandwiched contact thermocouple and split workpiece. Source: Ref 10

Contact thermocouples have proved effective in applications such as grinding (Ref 10) and brakes (Ref 11), but their accuracy remains questionable because of variable junction size and erratic transient response. Researchers have attempted to account for the thermal inertia of the thermocouple junction (Ref 10), but have not been completely successful.

Dynamic Thermocouples. In the dynamic thermocouple technique, sometimes called the Herbert-Gottwein technique, a thermocouple junction is formed at the sliding interface by the contacting bodies themselves. This method was originally developed to study contact temperatures at the interface between the cutting tool and the workpiece during metal-cutting operations (Ref 12). Later it was used to measure surface temperature in a variety of sliding contacts. Provided the two contacting materials are dissimilar and produce a well-characterized thermal emf as a function of temperature, the two rubbing materials can be used as part of a thermocouple circuit. An example of the use of the dynamic thermocouple technique is shown in Fig. 6 (Ref 13). In this application, a constantan ball served as one element of the thermocouple and a steel (iron) cylinder as the other. A measuring junction was formed wherever there was intimate contact between the ball and cylinder, and the measured temperature was the mean temperature at the contact interface (T_m).

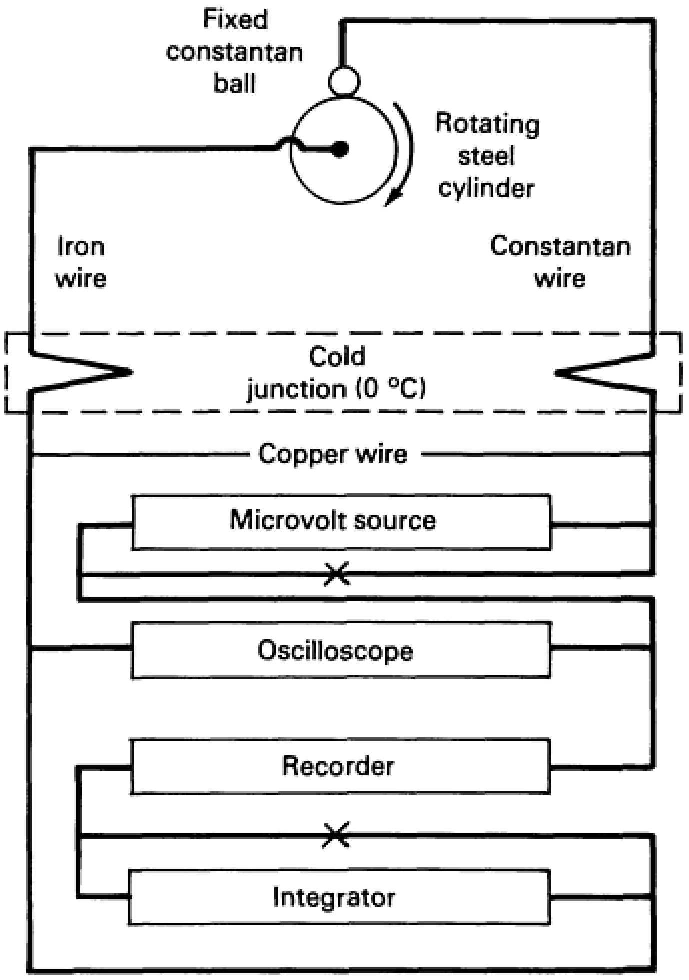


Fig. 6 Diagram of a dynamic thermocouple and monitoring circuitry. Source: Ref 13

In a modification of this technique, a single thin wire of constantan (or similar material) is embedded within the stationary sliding element, emerging at the contact surface. After the wire is insulated from the surrounding material with ceramic cement of epoxy, its end is ground smooth with the remainder of the sliding surface. In this way, a dynamic thermocouple junction will be formed whenever the end of the wire is in contact with the other surface, and the thermocouple location can be known with certainty (Ref 14).

Dynamic thermocouples have a very thin junction, consisting only of the actual contact zone. As a result, they can respond very rapidly to changes in surface temperature. Because they can respond to flash temperatures as well as to mean contact temperatures, dynamic thermocouples have been found to give higher measurements and faster transient response than embedded thermocouples (Ref 14). The measurements are often lower than theoretical predictions, however (Ref 13), and questions persist about the accuracy and meaning of the thermal emf produced by the thermocouple. The emf is produced by a weighted average of all temperatures across the sliding thermocouple junction, and that average may differ from the peak contact temperature (T_c), especially when the thermocouple junction contains more than one contact spot and the junction size changes with time (Ref 15). In addition, some electrical noise is frequently generated at the contact interface, and the thermocouple output must be discriminated from it.

Thin-Film Temperature Sensors. Over the past few years, microelectronics fabrication techniques, such as vapor deposition, have begun to be used to make surface temperature sensors. The advantage of such techniques is that very small sensors, with rapid response time, can be formed on surfaces.

One of the first vapor-deposited surface temperature sensors was a thermistor used to measure surface temperatures on gear teeth (Ref 16). The sensor consisted of a thin strip of titanium coated onto an alumina insulator on the surface of one of a pair of meshing teeth, as shown in Fig. 7 (Ref 16). The dimensions of the strip were approximately 500 μm wide by

2.5 mm long by 25 nm thick. The resistance of the titanium strip is sensitive to temperature, so monitoring the change in resistance allowed the transient surface temperature changes to be measured. By nature, the strip has a finite length and responds to all temperature changes along its length. Thus, it does not give a pointwise temperature measurement.

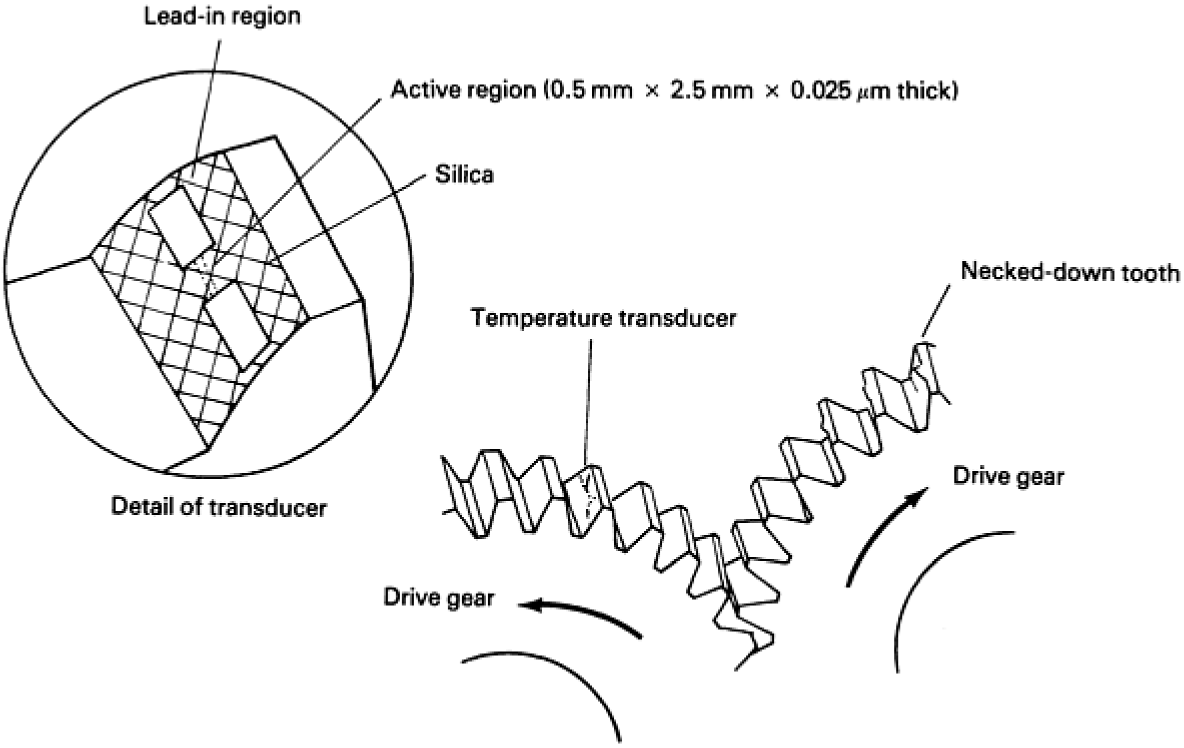


Fig. 7 Thin-film transducer (thermistor) for measuring surface temperature on gear teeth. Source: Ref 16

A more recent application of thin-film fabrication techniques in surface temperature measurement has been the development of thin-film thermocouples (TFTCs). Some of the earliest work on TFTCs described the use of vapor deposition to produce thermocouple pairs from thin films of nickel and iron, copper and iron, copper and nickel, copper and constantan, and chromel and alumel (Ref 17). Only recently have similar techniques been used successfully for measuring sliding surface temperatures (Ref 18). The production of a TFTC involves the deposition of thin films of two different metals, such as nickel and copper, sandwiched between thin layers of a hard, dielectric material such as Al_2O_3. A schematic cross section of a typical thin-film thermocouple is shown in Fig. 8 (Ref 18). The dielectric layer beneath the thermocouple junction is necessary to electrically insulate the TFTC device from the underlying metallic surface. Because of the softness of the thin metal films used for the thermocouple and connecting leads, a hard, protective layer must be deposited above the junction to limit damage to the TFTC device. The metal and dielectric films can be grown by physical vapor deposition techniques and can be less than 1 μm thick. Studies have shown that TFTC devices respond extremely rapidly (<1 μs) to a sudden temperature change and do not significantly disturb the heat flow from the sliding contacts (Ref 18). Such sensors can measure the temperature of the contact interface, especially when the protective layer is very thin, and the small size of the measuring junctions enables good indication of temperature at a specific point on the surface. If that point happens to be a flash temperature location, TFTCs can measure maximum contact temperature (T_c). As with all thermocouples or thermistors, however, thin-film devices cannot give a complete picture of surface temperature distribution, because that requires simultaneous measurements of temperature at a large number of points.

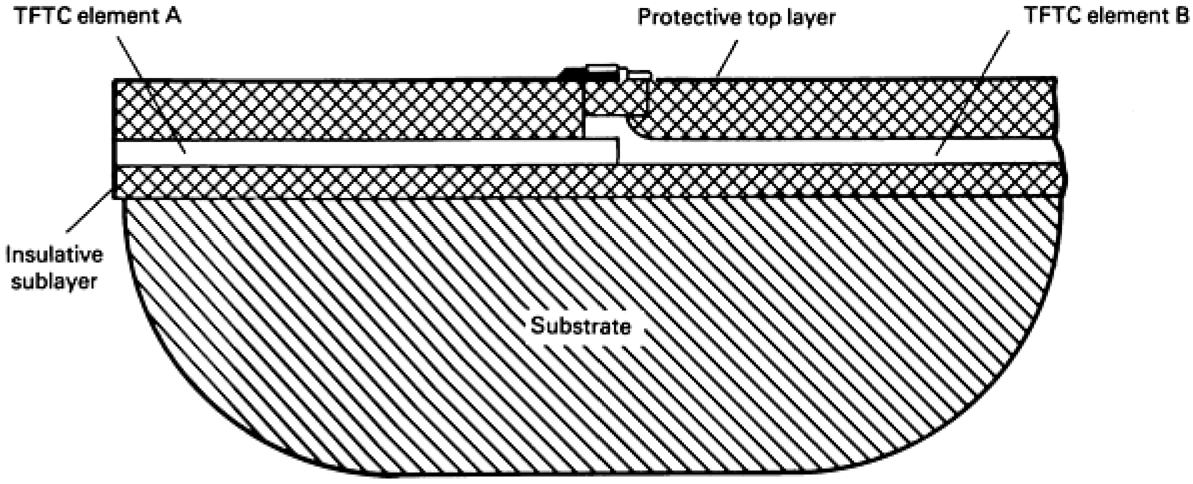


Fig. 8 Schematic cross section of a thin-film thermocouple. Source: Ref 18

Radiation Detection Techniques

Techniques involving the detection of thermal radiation have provided some of the most successful measurements of sliding contact temperatures. It is well known that any surface with a temperature above absolute zero is a natural radiator of thermal energy. The radiation emitted by the object is a unique function of its temperature, with the Stefan-Boltzmann law (Eq 1) showing that much more power is emitted by a body at high temperatures than at low ones:

$$\Phi = \epsilon \sigma T^4 A \qquad \textbf{(Eq 1)}$$

where Φ is the power (energy rate), T is the absolute temperature, σ is the Stefan-Boltzmann constant, A is the area of the heat source, and ϵ is the total emissivity of the surface emitting the radiation. ϵ is temperature dependent and is also very dependent on the characteristics of the surface. Some typical emissivity values at room temperature are listed in Table 1.

Table 1 Representative values of total emissivity of solid surfaces at 25 °C (77 °F)

Material	Total emissivity at 25 °C (77 °F)
Copper, polished	0.03
Copper, oxidized	0.5
Iron, polished	0.08
Iron, oxidized	0.8
Carbon	0.8

Source: Adapted from Ref 19

The radiation is composed of photons of many wavelengths, and, according to Planck's law, the monochromatic emissive power of a blackbody in a vacuum is:

$$W_{b,\lambda} = \frac{2\pi h C^2 \lambda^{-5}}{e^{Ch/k\lambda T} - 1} \qquad \textbf{(Eq 2)}$$

where λ is wavelength, C is the velocity of light in a vacuum, h and k are Planck and Boltzmann constants, respectively, and $w^{b,\lambda}$ is defined as the energy emitted per unit area at wavelength λ per unit wavelength in a small interval around λ. Integration of Eq 2 over all wavelengths leads to Eq 1 for the case of a blackbody (ϵ = 1). Planck's law (Eq 2) is plotted in

Fig. 9 for some representative temperatures (Ref 19). Note that $w_{b,\lambda}$ is very low at small and long wavelengths, so most emissive power is found at wavelengths in the range 1 μm < λ< 10 μm. For this reason, most successful attempts at measuring temperature by the detection of thermal radiation have concentrated on the infrared region of the spectrum (wavelengths of 0.75 to 500 μm). If the surface temperature *T* is high enough, radiation in the visible part of the spectrum (400 to 750 nm) can also be detected. Several different radiation measurement techniques have been successfully used to measure surface temperatures, including photography, pyrometry, thermal imaging, and photon detection.

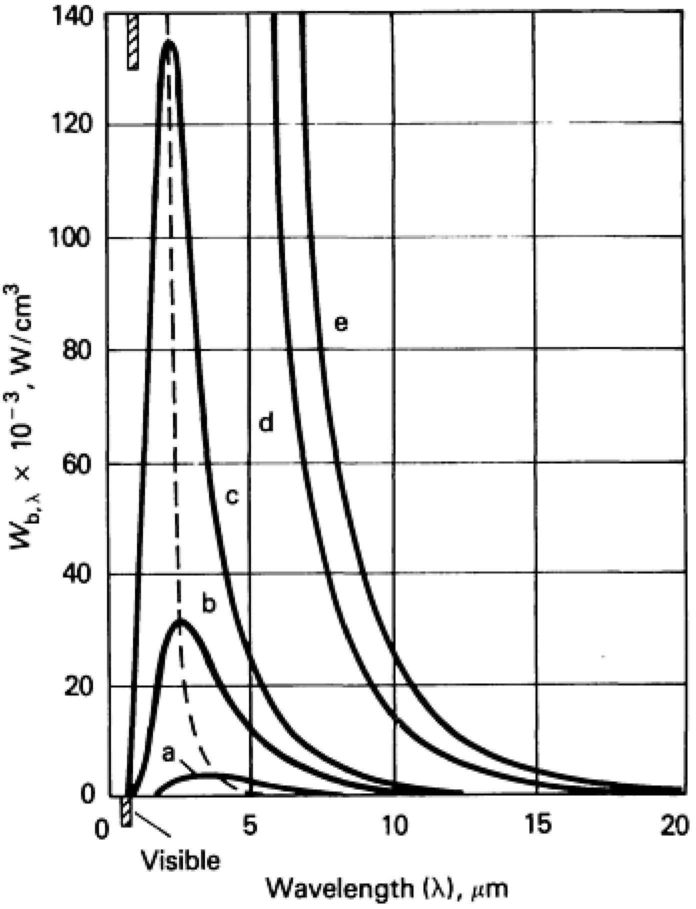


Fig. 9 Spectral radial emittance of a blackbody (Eq 1) at various temperatures. (a) 800 K. (b) 1200 K. (c) 1600 K. (d) 6000 K. (e) 10,000 K. Source: Ref 19

Optical and Infrared (IR) Photography. A photographic technique utilizing infrared-sensitive film was developed by Boothroyd in 1961 for studying the temperature distribution in metal cutting (Ref 20). Similar methods were used in later studies of surface temperatures in machining (Ref 21), as well as for sliding components such as brakes (Ref 5) and in pin-on-disk sliding (Ref 22). In most cases, the camera is focused on the moving body as it emerges from a sliding contact, but in recent years successful photographs have been made through a transparent window to a sapphire/metal or sapphire/ceramic contact (Ref 22). Sapphire is a useful material for such studies; its mechanical and thermal properties are similar to those of steel, and it is essentially transparent to radiation in the visible and near-infrared regions (Ref 23).

A photograph showing hot spots on a tool steel pin sliding at high speed on a sapphire disk is presented in Fig. 10 (Ref 22). Temperatures of the spots were estimated to range from 950 to 1200 °C (1740 to 2200 °F). The temperature distribution is best determined by measuring the optical density of the developed negative. The system must be calibrated to determine the density-temperature relationship of the film in the test configuration. This is usually accomplished by photographing specimens of the same material that have been heated to known temperatures and then comparing the optical density of the test film to that of the calibrated film. The same magnification and exposure time must be used in both the calibration and the test photographs. When careful procedures are used, this method can provide good indications of the surface temperature distribution on a sliding contact. In general, however, methods involving infrared-sensitive film require exposure times (5 s or more) that are longer than the duration of flash temperatures, so they probably do not

give a true indication of the highest temperature in a sliding contact. Shorter exposure times (less than 1 s) can be achieved with normal high-speed color film, as in Fig. 10, but such film operates with visible light and thus is restricted to detection of high temperatures, perhaps 800 °C (1470 °F) or higher. Infrared film can respond to mean contact temperatures as low as 300 °C (570 °F), but even those temperatures occur only in rather severe sliding situations. An alternative is to use a modern infrared camera, which is essentially a scanning infrared detector.

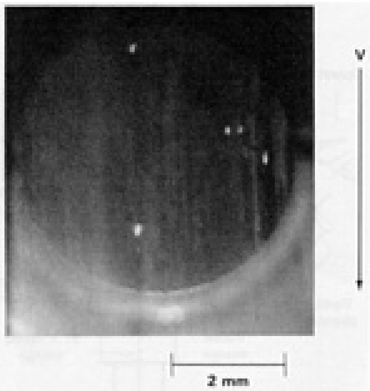


Fig. 10 Hot spots on a steel pin (6 mm, or 0.24 in., diam) sliding at 2 m/s (6.5 ft/s) on a sapphire disk with a load of 26 N (5.8 lbf). Photograph taken with back lighting after 25 min of sliding. Exposure time, 1 s. Source: Ref 22

Infrared detectors have been widely used and improved since Parker and Marshall (Ref 24) employed an optical pyrometer to measure the surface temperature of a railroad wheel as it emerged from a brake shoe. Early pyrometers used the eye as a detector to match the brightness of the subject body with that of a standard lamp incorporated in the instrument. Improved models that employed a photoelectric detector were later developed. The detector essentially integrates Eq 2 over all wavelengths within its spectral range and over the surface area viewed by the detector. Because the temperature is generally not constant over the field of view, the detector output is a function of the average temperature over the area.

To improve the accuracy of the temperature measurement and to approach a point measurement, most modern detectors are equipped with optics that limit the field to view to a small spot size, perhaps on the order of 100 to 500 μm diam. The result is an IR radiometric microscope. An example of which is shown schematically in Fig. 11 (Ref 25). This microscope uses a liquid-nitrogen-cooled detector made of indium antimonide (InSb), which has a spectral band of 2 to 5.6 μm. Infrared radiometric microscopes can measure transient temperature changes at a rate of up to 20 kHz or greater. They have been used effectively both with metallic components, where the detector is focused on a spot just emerging from the contact zone (Ref 25), and with transparent sapphire components, where the detector is focused through the sapphire onto the contact zone between the sapphire and the metal (Ref 26).

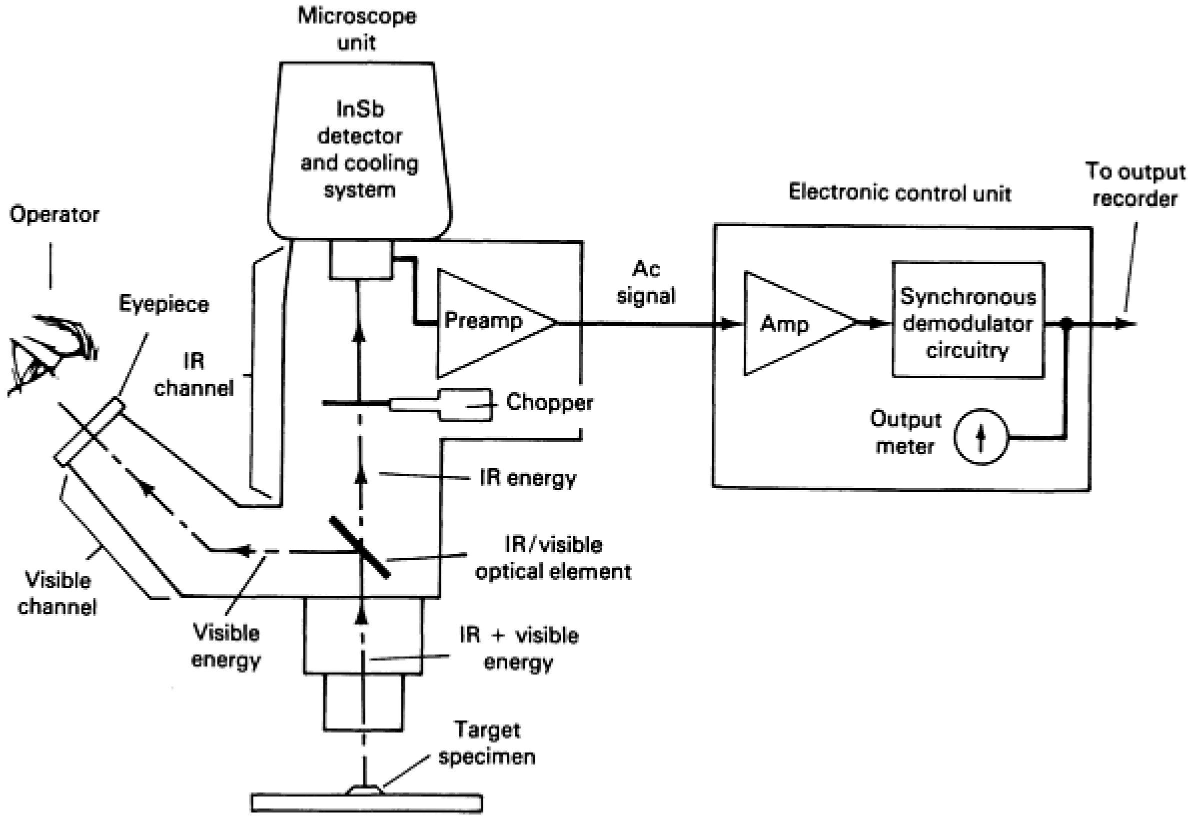


Fig. 11 Schematic diagram of an IR radiometric microscope. Source: Ref 25

By limiting the field of view to a single small spot, the IR microscope can miss many contact events occurring at other spots within the area of contact. To overcome this limitation, in recent years investigators have been using a scanning IR camera to study sliding surface temperatures (Ref 23, 27, 28). A scanning IR camera, or IR microimager, has a detector similar to that shown in Fig. 11, but the detector is optically scanned over the contact surface in either of two modes: line scan or area scan. In the line scan mode, a fixed line, perhaps several millimeters in length, is scanned continuously at a rate of up to 2500 line scans per second. In the area scan mode, the rotation of a prism advances the line for each scan to produce a field typically consisting of 100 scan lines. The output is a video voltage that is a function of the IR radiation detected at that instant. The output from a complete area scan for the case of a silicon nitride pin sliding against a sapphire disk is shown in Fig. 12 (Ref 23).

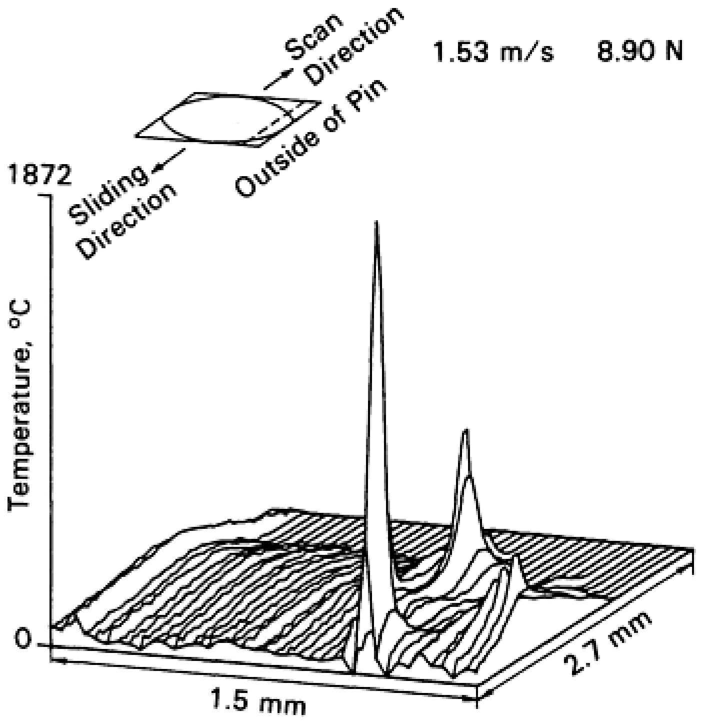


Fig. 12 Temperature distribution on the surface of a silicon nitride pin in sliding contact with a sapphire disk at a velocity of 1.5 m/s (5 ft/s) and a normal load of 8.9 N (2 lbf). Pin bulk temperature, 96 °C (205 °F). Area scan mode with IR scanning camera. Source: Ref 23

It should be noted that, even at a scan rate of 2500 lines per second, it takes 40 ms to complete an area scan composed of 100 lines. This is considerably longer than the duration of flash temperatures, so it is unlikely that a field plot similar to that in Fig. 12 is fully representative of the surface temperature distribution at any instant. However, it does give a good indication of the contact conditions within the target area and the approximate temperatures reached at the hot spots. A better indication of transient temperatures at a given point can be achieved in the line scan mode, by continuously sweeping over the same line. Even in that mode, however, the transient times of the temperature fluctuations have been found to be less than the time required to complete a single line scan (Ref 23), and the flash temperature intervals may, in fact, be less than the 5 μs or so between consecutive temperature measurements on the same scan line. Thus, measured contact temperatures may be less than actual flash temperatures, particularly if the hot spot is smaller than the spot size of the detector and is very short-lived.

Methods have been devised to correct for the instantaneous temperature averaging that occurs within an IR detector (Ref 23), but such techniques are only approximate. As stated earlier, an IR detector essentially integrates Eq 2, multiplied by the emissivity, over the area of the detector's target spot and within its spectral range to obtain the equivalent of Eq 1. If a small hot spot whose temperature is desired is contained within a larger target spot, the area of the spot must be known so that its contribution to the summed detector output can be determined. Because hot-spot areas are usually not known with certainty, the hot-spot temperature is inaccurately determined. A better technique, utilizing two separate detectors, was recently devised by Bair *et al.* (Ref 29). If the emitted radiation is split between the two detectors and a different band-pass filter is placed in front of each detector, different values of radiated power will be measured at each of the two wavelengths, and each will be a function of two variables: hot-spot area and temperature. The ratio of detected power at the two wavelengths can be used to determine the maximum temperature within the field of view (Ref 29). The hot-spot area can also be determined once its temperature has been calculated. The optical setup for this method is shown in Fig. 13, and a typical results is shown in Fig. 14, for a sapphire read/write slider against a thin-film magnetic disk (Ref 29). This method appears to hold much promise for future development.

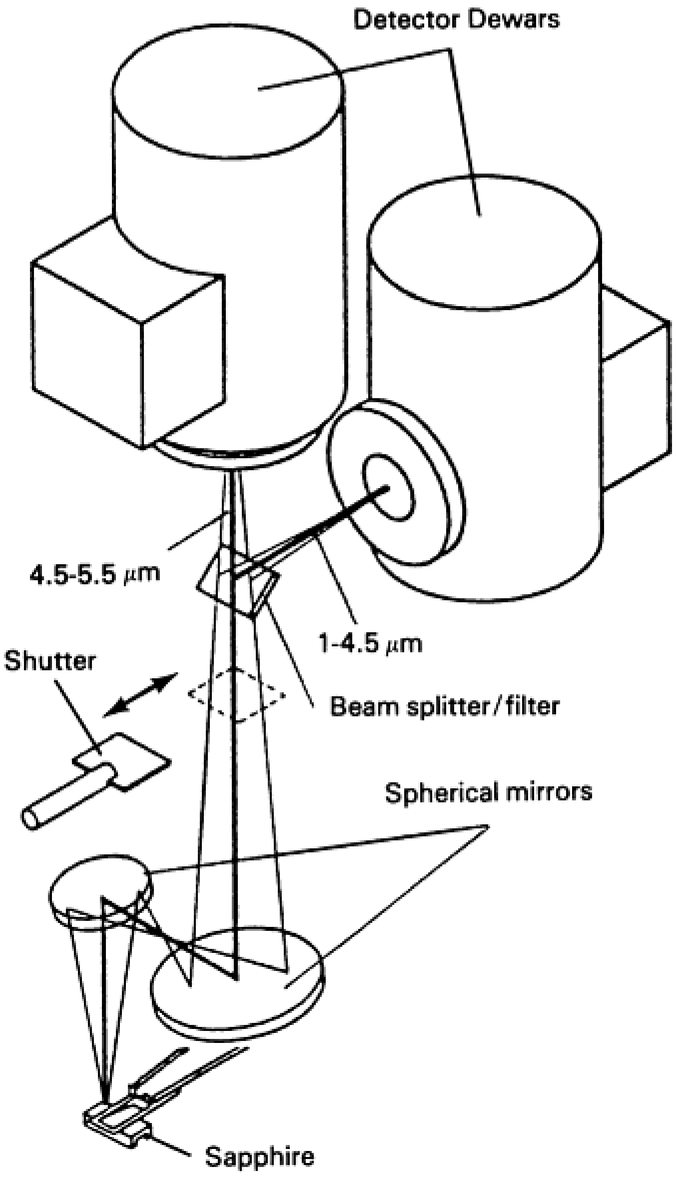


Fig. 13 Optical arrangement for an IR temperature measurement system with two detectors. Source: Ref 29

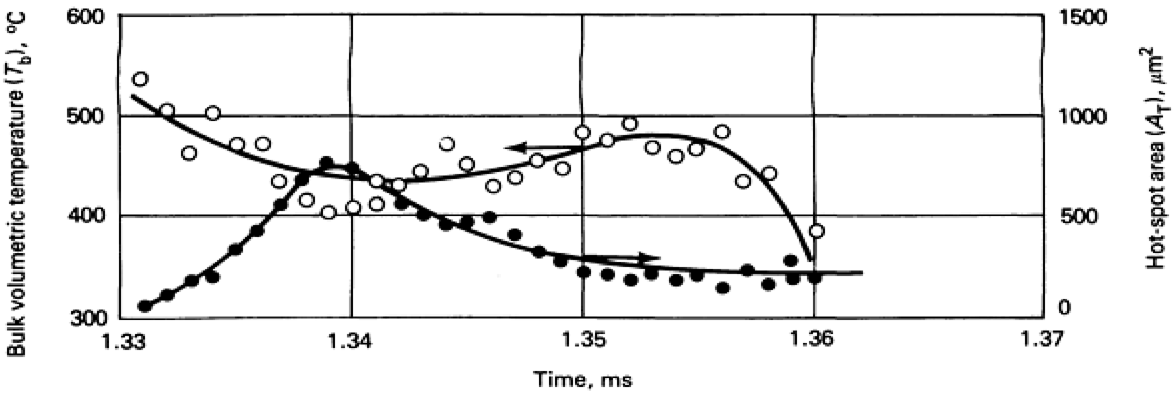


Fig. 14 Hot-spot temperature and area measured using an IR temperature measurement system with two detectors. Source: Ref 29

Uncertainty about the emissivity of contacting surfaces during the sliding process can lead to inaccurate temperature determination with any of the IR techniques. It is apparent from Table 1 that the total emissivity of a metallic (or nonmetallic) surface can vary considerably owing to oxidation, wear, or other changes in surface characteristics. To obtain an accurate temperature reading from a radiating surface, an accurate value of emissivity must be known at that temperature. This can be accomplished by carefully determining the emissivity of reference surfaces similar to the contacting surfaces at temperatures throughout the range of interest. There are also methods for measuring the emissivity and transmissivity of any lubricant between the surfaces (Ref 28). Despite these procedures, emissivity remains an accuracy-limiting variable in many IR measurements of sliding surface temperatures, especially when the emissivity changes during the sliding process (Ref 27). These difficulties have been partially overcome by the technique of Bair *et al.* (Ref 29), in which the calculated temperature is independent of the emissivity, provided the spectral distribution of emissivity remains unchanged.

Photon Collection. As an alternative to IR detectors, which generally respond too slowly to accurately measure flash temperatures in rapidly moving hot spots, a surface temperature measurement method was recently developed that uses a photomultiplier to collect photos emitted by hot contact spot (Ref 30). An experimental setup using this technique is shown in Fig. 15. Use of a photomultiplier with sensitivity to wavelengths in the 500 to 900 nm range allows detection of photons emitted from a hot contact between a moving surface and a transparent (sapphire) slider. Contact temperatures generally need to be at least 400 to 500 °C (750 to 930 °F) in order to generate photons with enough energy to be detected by the photomultiplier (see Fig. 9), but the response time of the photomultiplier is very rapid (less than 30 ns). Therefore, the technique can be used for detecting flash temperatures of very hot duration (2 μs or less) (Ref 30). However, it is not too useful for measuring mean contact temperatures, which are generally lower than 500 °C (930 °F). A further restriction is that the sliding test must be run in complete darkness to eliminate noise, which can dominate the output signal. This method is also subject to a limitation similar to that of most IR detectors; that is, the area of the spot emitting the photons must be known in order to accurately determine the temperature of the spot. Calibration of the relationship between output voltage and temperature can be accomplished using reference surfaces of the contacting materials at known temperatures (Ref 30).

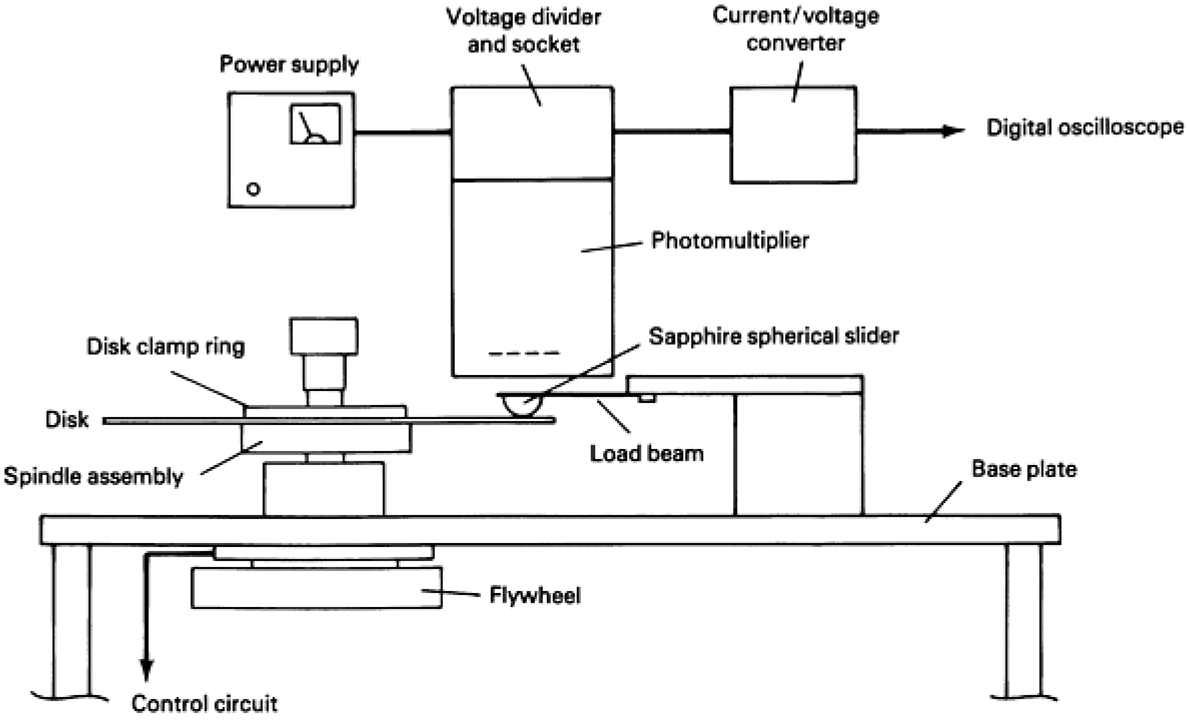


Fig. 15 Experimental setup for surface temperature measurement using photon collection. Source: Ref 30

Conclusion

The techniques described in this article have led to much useful information about the temperatures generated by frictional heating. Although each of the techniques is a useful engineering or research tool, each is subjects to some

limitations; therefore,no single surface temperature method can be used for all applications. Some techniques, such as thermocouples, give a measure of temperature at a single point, while others, such as IR imaging, give a field measurement. Response time varies considerably, from less than 50 ns for photon detection, to 1 μs or less for thin-film thermocouples, to several seconds for IR photography, to nearly a minute for metallographic techniques. Some methods are more useful as research tools than in studies using actual machine components. It is likely that a complete study of sliding surface temperatures in any real application would involve the coordinated use of several of the measurement techniques described above

Before beginning a surface temperature measurement task, the engineer should remember that the real area of contact in actual mechanical components is quite inaccessible and is made up of small spots that may remain in contact for a very short time. During that short contact time, the spots can get much hotter than the rest of the material within the nominal contact zone, attaining very high flash temperatures. Despite the progress that has been made in surface temperature measurement, it is difficult to accurately measure localized flash temperatures in actual mechanical components; this has led many investigators to rely on analytical methods for contact temperature determination. Experimental temperature measurements are often used in conjunction with analysis in order to provide necessary boundary condition information or to verify analytical predictions.

References

1. H. Block, Theoretical Study of Temperature Rise at Surfaces of Actual Contact under Oiliness Lubricating Conditions, *Proceedings: General Discussion on Lubrication and Lubricants,* Vol 2, Institution of Mechanical Engineers, 1937, p 222-235
2. P.K. Wright, Correlation of Tempering Effects with Temperature Distribution in Steel Cutting Tools, *J. Eng. Ind.,* Vol 100, 1978, p 131-136
3. P.K. Wright and E.M. Trent, Metallographic Methods of Determining Temperature Gradients in Cutting Tools, *J. Iron Steel Inst.,* Vol 211, 1973, p 364-388
4. R.P. Reed, Thermoelectric Thermometry: A Functional Model, *Temperature--Its Measurement and Control in Science and Industry,* J.F. Schooley, Ed., American Institute of Physics, Vol 5, 1982 p 915-922
5. J.J. Santini and F.E. Kennedy, An Experimental Investigation of Surface Temperatures and Wear in Disk Brakes, *Lubr. Eng.,* Vol 31, 1975, p 402-417
6. F.F. Ling and T.E. Simkins, Measurement of Pointwise Junction Condition of Temperature at the Interface of Two Bodies in Sliding Contact, *J. Basic Eng.,* Vol 85, 1963, p 481-487
7. S. Suzuki and F.E. Kennedy, Friction and Temperature at Head-Disk Interface in Contact Start-Stop Tests, *Tribol. Mech. Magn. Storage Sys.,* Vol 15, 1988, p 30-36
8. G.A. Berry and J.R. Barber, The Division of Frictional Heat--A Guide to the Nature of Sliding Contact, *J. Tribology (Trans. ASME),* Vol 106, 1984, p 405-415
9. A.V. Yakimov, Investigation of Temperature in the Grinding Zone, *Russ. Eng. J.,* Vol 8, 1964, p 51
10. A.Y.C. Nee and A.O. Tay, On the Measurement of Surface Grinding Temperature, *Int. J. Mach. Tools Des. Res.,* Vol 21, 1981 p 279-291
11. V.A. Balakin, Heat Flow Distribution and Combined Heat-Mass Transfer Processes at the Contact Interface of a Friction Pair, *J. Eng. Phys.,* Vol 40, 1981, p 660-665
12. H. Shore, Thermoelectric Measurement of Cutting Tool Temperature, *J. Wash. Acad. Sci.,* Vol 15, 1925, p 85-88
13. M.J. Furey, Surface Temperatures in Sliding Contact, *ASLE Trans.,* Vol 7, 1964, p 133-146
14. F.E. Kennedy, Single Pass Rub Phenomena--Analysis and Experiment, *J. Lubr. Technol. (Trans. ASME),* Vol 104, 1982, p 582-588
15. H.H. Shu, E.W. Gaylord, and W.F. Hughes, The Relation between the Rubbing Interface Temperature Distribution and Dynamic Thermocouple Temperature, *J. Basic Eng.,* Vol 86, 1964, p 417-422
16. J.W. Kannel and J.C. Bell, A Method for Estimation of Temperatures in Lubricated Rolling-Sliding Gear or Bearing Elastohydrodynamic Contacts, *Institution of Mechanical Engineers EHD Symposium,* 1972, p 118-130
17. R. Marshall, L. Atlas, and T. Putner, The Preparation and Performance of Thin Thermocouples, *J. Sci.*

Instrum., Vol 43, 1966, p 144

18. X. Tian, F.E. Kennedy, J.J. Deacutis, and A.K. Henning, The Development and Use of Thin Film Thermocouples for contact Temperature Measurement, *Tribol. Trans.*, in press (1992)
19. R.E. Bedford, Blackbody Radiation, *Encyclopedia of Physics, 2nd ed.*, R.G. Lerner and G.L. Trigg, Ed., 1991, p 104-105
20. G. Boothroyd, Photographic Technique for the Determination of Metal Cutting Temperatures, *Br. J. Appl. Phys.*, Vol 12, 1961, p 238
21. S. Jeelani, Measurement of Temperature Distribution in Machining Using IR Photography, *Wear*, Vol 68, 1981, p 191-202
22. T.F.J. Quinn and W.O. Winer, The Thermal Aspects of Oxidational Wear, *Wear*, Vol 102, 1985, p 67-80
23. J.A. Griffioen, S. Bair, and W.O. Winer, Infrared Surface Temperature Measurements in a Sliding Ceramic-Ceramic Contact, *Mechanisms and Surface Distress*, D. Dowson *et al.*, Ed., Butterworths, 1986, p 238-245
24. R.C. Parker and P.R. Marshall, The Measurement of the Temperature of Sliding Surfaces With Particular Reference to Railway Blocks, *Proc. Inst. Mech. Eng.*, Vol 158, 1948, p 209-229
25. D.P. Townsend and L.S. Akin, "Analytical and Experimental Spur Gear Tooth Temperature as Affected by Operating Variables," Paper 80-C2/DET-34, presented at International Power Transmission and Gear Conference (San Francisco), Aug 1980, American Society of Mechanical Engineers
26. A. Floquet and D. Play, Contact Temperatures in Dry Bearings--Three Dimensional Theory and Verification, *J. Tribology (Trans. ASME)*, vol 103, 1981, p 243-252
27. M.A. Meinders, D.F. Wilcock, and W.O. Winer, Infrared Temperature Measurements of a Reciprocating Seal Test, *Proceedings of the 9th Leeds-Lyson Conference*, IPC Science and Technology Press, 1982, p 321-328
28. B. Tournerie, D. Reungoat, and J. Frene Temperature Measurements by Infrared Thermography in the Interface of a Radial Face Seal, *J. Tribology (Trans. ASME)*, Vol 113, 1991, p 571-576
29. S. Bair, I. Green, and B. Bhushan, Measurements of Asperity Temperatures of a Read/Write Head Slider Bearing in Hard Magnetic Recording Disks, *J. Tribology (Trans. ASME)*, Vol 113, 1991, p 547-554
30. S. Suzuki and F.E. Kennedy, The Detection of Flash Temperatures in a Sliding Contact by the Method of Tribo-Induced Thermoluminescence, *J. Tribology (Trans. ASME)*, Vol 113, 1991, p 120-127

Surface Chemical Analysis

Austin L. Grogan and Stephen L. Rice, University of Central Florida; Mark Davidson, University of Florida

Introduction

THE TECHNIQUES AND METHODOLOGIES of chemical analysis that can be applied to problems in friction, lubrication, and wear are the focus of this article. The analytical techniques discussed are x-ray photoelectron spectroscopy, ion-scattering spectrometry, Auger electron spectroscopy, electron-stimulated desorption, and secondary ion mass spectrometry.

These techniques yield chemical information that is specific for the surface and near-surface regions of solid-state materials. They are therefore well suited to the types of problems often encountered in friction, lubrication, and wear. These techniques can also be used to obtain chemical information about bulk regions of a material when ion etching is employed to remove surface layers of the material. A brief description of infrared spectroscopy is also provided, because it can be utilized in conjunction with the surface analytical techniques described below.

X-Ray Photoelectron Spectroscopy (XPS)

XPS is a near-surface sensitive analytical technique based on the photoelectric effect, which was first indirectly observed by Hertz in 1887 (Ref 1). He noted that it was easier to generate a spark from an electrode when the electrode was illuminated with ultraviolet (UV) light. Thompson (Ref 2) was able to show that the results of Hertz are due to the emission of electrons that occurs with exposure to light, and Einstein (Ref 3) used these results to develop the theory of quantum radiation, upon which modern XPS is based.

The origins of XPS date to the late 1960s, as summarized in Ref 4 and 5, which describe the theory and applications in some detail. Because the energy of an emitted electron indicates both the species and the oxidation state of that species, the term electron spectroscopy for chemical analysis (ESCA) was coined. It has been suggested by Carlson (Ref 6) that the term ESCA be used in reference to all photoelectron spectroscopies (for example, XPS and ultraviolet photoelectron spectroscopy) and that XPS be used to refer to photoelectron spectroscopy that is excited by x-rays. This convention will be followed in this article.

Since the publication of Ref 4 and 5, the field of ESCA has expanded from an experimentally difficult technique that was routinely performed in only a few laboratories to a routine analytical tool. This has been due, in large part, to the development of vacuum technology and the availability of standard vacuum pumps and equipment. Presently, XPS has advanced to the point where it is now being considered for use in process control schemes (Ref 7). Commercial x-ray photoelectron spectrometers are now routinely available from numerous vendors with a variety of options, which often include other complementary surface analytical techniques in the same vacuum chamber.

XPS and Auger electron spectroscopy (AES) have now become the most routinely used surface analytical techniques in many fields. The probing depth of XPS ranges from 1 to 5 nm, and the detection limit typically ranges from 1 part per 100 to 1 part per 500, with a limit of 1 part per 1000 being approached under the most favorable of conditions. XPS has proved to be useful in such widely diverse areas as semiconductor technology (Ref 8), catalysis (Ref 9), electrochemistry (Ref 10), and materials research (Ref 11, 12, 13, and 14). Reference 15 provides the early history of ESCA. A more recent history is provided in the introduction in Ref 6. Examples of previous applications of XPS to problems in friction, lubrication, and wear are provided in Ref 16, 17, 18.

XPS Fundamentals. ESCA is based on the photoelectric effect, which is simply the ejection of an electron from an atom upon absorption of a photon. This process is shown in Fig. 1. If the energy of the photon is greater than the binding energy of the electron, then the electron is ejected from the atom. The energy of the electron is given by an energy balance (Ref 16):

$$KE = h\nu - BE - \phi \qquad \textbf{(Eq 1)}$$

where KE is the kinetic energy of the ejected electron, $h\nu$ is the energy of the photon, BE is the energy of the electron in the atom, and ϕ is the work function of the spectrometer (the energy required for an electron to escape its image charge).

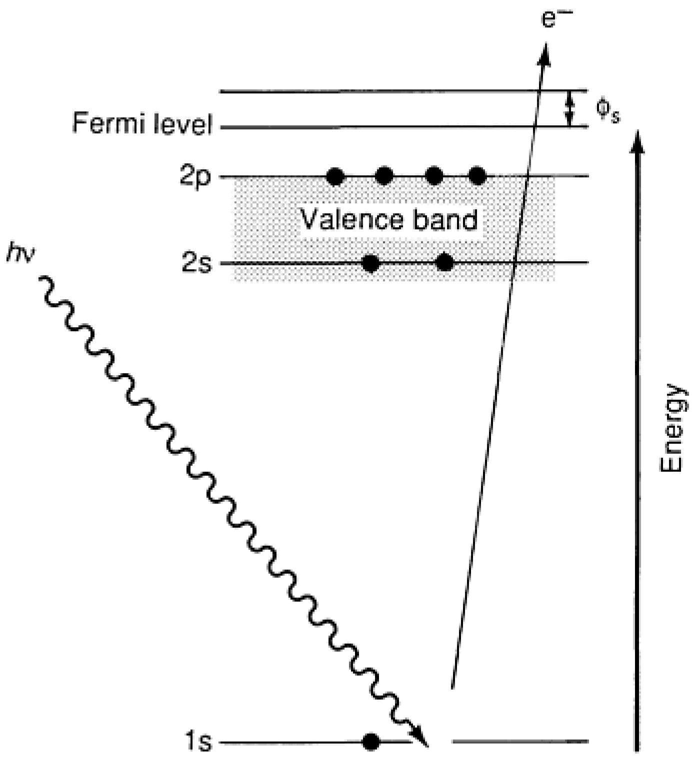


Fig. 1 Photoelectric effect

By supplying a flux of photons of known energy and measuring the energy of the ejected electrons, the binding energies of the electrons can be determined. The binding energies are indicative of the atoms from which the electrons came, and the flux of ejected electrons with a given binding energy is proportional to the number of atoms in the analyzed region. This information can be used to both identify and quantify the species present in the material under analysis. In addition, the binding energies of the core electrons can be shifted by changes in oxidation state. This shift is due to differing levels of screening of the core electrons by the valence electrons. By looking at these relatively small changes (0.1 to 10 eV) the chemical state of the analyzed species can be determined (Ref 4, 5, 14).

An example of this effect is shown in Fig. 2. These spectra were obtained from a TiO_2 (001) surface that had been sputtered (further described in the section "Auger Electron Spectroscopy"), annealed in ultrahigh vacuum (UHV), and then annealed in oxygen. The Ti 2p peaks widen and shift as the sputtering process selectively removes oxygen, and peaks from both Ti_2O_3 and TiO appear.

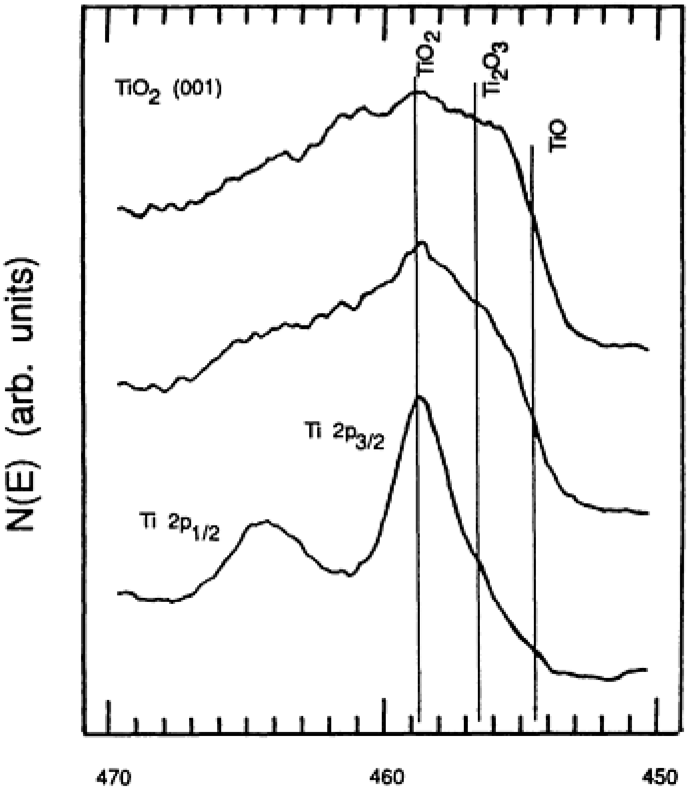


Fig. 2 XPS spectra from TiO_2 (001) surface showing Ti 2p peaks taken from (a) sputtered surface; (b) sputtered surface after annealing under vacuum at 400 °C (750 °F) for 1 h; (c) sputtered surface after annealing O_2 at 400 °C (750 °F) at 1.3 × 10^{-4} Pa (10^{-6} torr) for 1 h. N(E) = $N_{(E)}$. Source: Ref 19

The quantum mechanics of the photoelectric effect are much more complex than the simple theoretical explanation given here. In actuality,the energy balance may not give exactly the *BE* of the electron, because it assumes that the ion left behind is in its ground state (Ref 20). In practice, this occasionally gives rise to small background peaks called shake-up lines (Ref 21). For a thorough discussion of the theoretical basis for ESCA, see Ref 5, 6, 8, 20, and 22.

XPS Equipment. The experimental setup for XPS can be divided into four main components. There must be a source of x-rays, an electron energy analyzer, a UHV chamber in which to house them, and a sample to analyze. The vacuum chamber and associated pumps are standard items and are not discussed here, but the limitations and capabilities of the most popular options for each of the other three parts are described.

XPS X-Rays Sources. Conventional x-ray tubes that utilize either an MgKα or AlKα source (Fig. 3) represent the principal source of x-rays used in XPS. For most surface studies, the MgKα source is optimal, because of its higher intensity and narrower line width (Table 1), compared with other sources. The AlKα source is useful for observing higher binding energy lines, such as the Ge 2p lines at binding energies between 1230 and 1300 eV (Ref 21). Dual anode sources that give the researcher the option of the higher energy photons, if needed, are available. In addition, other anode materials are available with energies up to 8048 eV (CuKα), but for routine analysis, either the Mg or Al anode generally suffices.

Table 1 Energy, line width, and intensity values for several anode materials

Anode	Energy, eV	Line width, eV	Intensity
Mg	1253.6	0.7	1
Al	1486.6	0.85	0.5
Zr	2042.4	1.7	0.05
Cu	8048	. . .	. . .

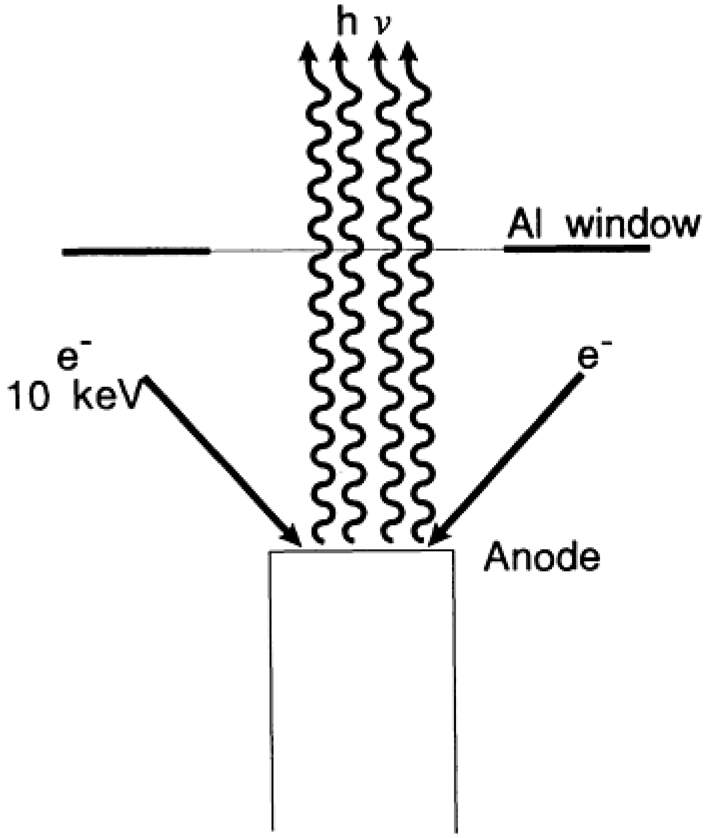


Fig. 3 Typical x-ray source, where electrons are accelerated to about 10 keV to the anode material, which then emits photons with energy characteristic of the anode material

As noted by Gardella (Ref 23) and Hercules (Ref 24), the conventional x-ray sources have great limitations with respect to energy and spatial resolution. The maximum energy resolution limitation of energy analyzers used in XPS systems is typically much less than the 0.7 eV natural line width of the MgKα x-ray source. Therefore, the energy resolution of the XPS data is limited by the x-ray source.

In addition, the x-ray spectrum emitted from the x-ray source contains minor x-ray lines (satellites) and a continuum background radiation called brehmsstrahlung. This prompted Fellner-Feldegg *et al.* (Ref 25) to develop a monochromatized x-ray source. These sources are now commercially available and have been used to measure line widths down to 0.34 eV. The monochromators have the additional advantage of providing focusing capabilities down to 150 μm. This allows the spatial mapping of surfaces with the power of chemical state determination provided by XPS (Ref 23).

Future developments in x-ray sources include high-flux rotating anode types that allow high energy resolution spectra to be obtained in 1 s. New anode materials will provide higher energy and narrower lines, and spots sizes as small as 1 μm (Ref 23).

Energy Analyzers. Two types of energy analyzers are often used in XPS spectrometers: the cylindrical mirror analyzer (CMA) and the hemispherical analyzer, both of which are the energy-dispersive type. The CMA consists of two coaxial cylinders with a potential placed between them (Fig. 4). The charged particles enter the region between the cylinders and are deflected by the electric field. The amount of deflection depends on the energy of the particle. Only those particles in a very narrow band can reach the detector. To make the spectrometer capable of the resolution necessary for XPS, two CMAs are placed in series (Ref 26). This configuration is called a double-pass CMA. The hemispherical analyzer is similar in operation to the CMA, but its geometry consists of two concentric hemispheres instead of cylinders.

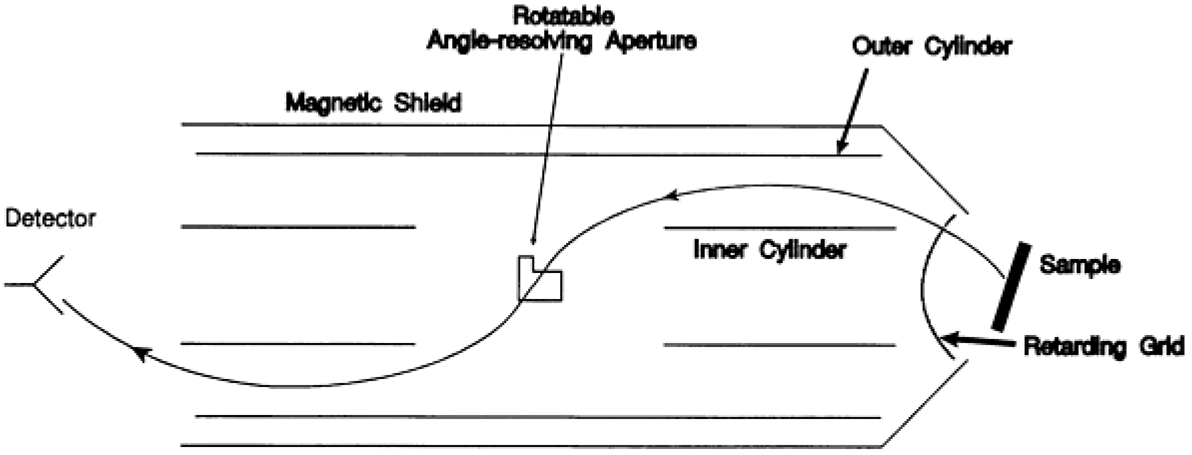


Fig. 4 Double-pass CMA showing charged particle path. Note that detection angle from sample can be varied by rotating central angle resolving aperture.

Both types of analyzers have advantages and disadvantages. The hemisphere generally has a higher throughput than the CMA for a given resolution. This can be important for XPS systems, because XPS signals typically have a low *S/N* ratio. Because the hemispherical analyzer has a higher throughput and, therefore, a higher signal for the same resolution, the XPS spectrum collected with the hemisphere will generally have a higher *S/N* ratio than that collected with the CMA for the same collection time. A spectrum with the same *S/N* can be collected with the CMA, but it will generally require longer collection times. However, the CMA is a more versatile instrument. Although angle-resolved XPS can easily be performed with the hemispherical geometry, the CMA can be equipped with an angular resolving aperture, which, unlike the hemispherical analyzer, allows the independent variation of excitation and detection angles. This is important for angle-resolved XPS and AES, which yield depth compositions nondestructively (Ref 14, 27), and for ion-scattering spectroscopy. The CMA is also better suited to collect AES data than the hemisphere. In general, the hemisphere is a better choice for a dedicate XPS instrument and the CMA is best for a multiple-technique instrument, especially if angle resolved experiments are desired. The CMA analyzer will be discussed further in the section Auger Electron Spectroscopy (AES) in this article.

Sample and Preparation. XPS requires a UHV environment to prevent arcing in the x-ray source, to prevent the photoelectron from being scattered from gas phase molecules, and to reduce accumulation of surface contamination during analysis. These requirements place some limitations on sample types, as well as preparation. The samples must have a low vapor pressure and be free from any contaminants, such as oil and fingerprints. As previously stated, the analysis is limited to the near-surface (1 to 5 nm) region of the sample. The addition of an ion gun allows sputter depth profiling, but the depth is limited to a few microns for a reasonable experimental time scales.

Preparation of the sample can be accomplished in several ways, depending on the nature of the sample. A common procedure for cleaning and mounting is as follows: The sample is ultrasonically cleaned in a series of solvents. Typically, the deionized water is first used to remove inorganic constituents, followed by cleaning in toluene, acetone, trichloroethylene, and then acetone and ethanol. This cleaning procedure sufficiently removes most contamination in the surface and near-surface regions.

If the desired, the final removal of remaining contamination can be accomplished by ion sputtering and/or annealing in the vacuum system. If the removal of surface oxide layers is required, such as in many single-crystal applications, the specimen can be etched with an appropriate acid. Buffered hydrofluoric solutions are commercially available for this purpose.

There are several methods for mounting samples such as powders. In one method, a sample of the powder is pressed between two soft metal foils, such as tin or indium. The foils are then separated and one of them is mounted for analysis. The authors have had mixed results with this method, because the substrate often appears in the spectra obtained, and the diversity in sizes of the "islands" of sample powder can lead to different charging levels, causing multiple peaks in the spectrum. The best results have been obtained by using a small aluminum cup to hold the powder. The powder is pressed into the cup with a pellet press and then the cup and powder pellet are mounted. This method is particularly useful when

the sample is a pellet to begin with and the average concentrations throughout the pellet are desired. The pellet can be ground into a fine powder and then pressed into an aluminum cup for mounting. Care should be taken to grind slowly, so as not to produce excessive heat.

XPS Spectrum. A sample spectrum obtained from a manganese dioxide powder is shown in Fig. 5. The horizontal scale is labeled as binding energy, and it should be noted that the spectrum is displayed with the binding energy increasing from the right to left. This is a customary way to display XPS spectra, because the data are taken as a function of kinetic energy, which increases from left to right.

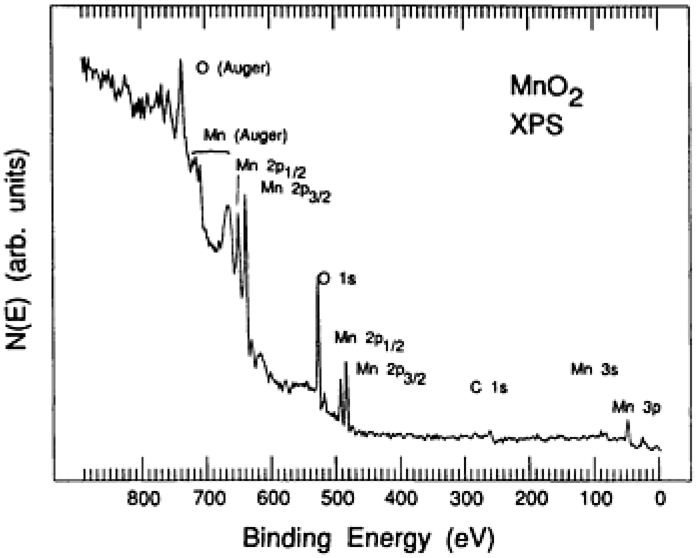


Fig. 5 Sample XPS spectrum of MnO_2 powder showing valence and core photoelectron peaks, as well as Auger peaks. N(E) = $N_{(E)}$

Several features are evident in the spectrum. The most noticeable trend is the overall increase in the background, from low to high binding energy. This increase is due to inelastic scattering of photoelectrons as they move through the sample. As the electrons scatter, they lose kinetic energy to the solid and thus contribute to the background towards low kinetic energy (high binding energy). This is evidenced by the sudden increases in the background at the photoemission lines. Below the kinetic energy level at which the electrons are emitted, there are more electrons that can inelastically scatter. Thus, the background increases suddenly and in proportion to the magnitude of the photoelectron emission.

The most important features in the spectrum are the peaks that are due to the photoemission of the core electrons. The energy and magnitude of these peaks are the basis for the qualitative and quantitative analyses of the sample, which are discussed in detail in next section. Also present in the spectrum are peaks that are due to Auger electrons created during the relaxation of the excited (core hole) that remains after the photoemission process.

The Auger process will be discussed in more detail in the section Auger Electron Spectroscopy (AES) , but it should be noted that the kinetic energy of the Auger electron is independent of the energy of the original excitation. This fact can be exploited in avoiding interferences between photoelectron lines and AES lines in the XPS spectrum.

If a dual anode x-ray source is available, the kinetic energy of the photoelectron peaks can be shifted by changing x-ray energies, but the kinetic energy of the Auger lines will remain constant and the interference will be removed. There are also peaks in the region of binding energies near the Fermi level (defined as zero binding energy) that are due to valence electrons. These features can be of considerable use in surface studies, although they are generally small in magnitude, because of the relatively low photon flux from typical x-ray tubes and their low probability of interacting with valence electrons. Useful valence band spectra can be obtained using UV light as an excitation source. This low-energy radiation couples well with valence-level electrons. Therefore, valence-level, rather than core-level, electrons are emitted. This technique, referred to as ultraviolet photoelectrons spectroscopy (UPS), is a powerful analytical technique, but the spectra are typically complicated to interpret because of Molecular orbital formation and chemical state effects. Therefore, UPS is not used on a routine basis for most applications.

Two small and less important features it the spectrum are x-ray satellites and shake-up satellite. X-ray satellites are a result of excitation of the photoelectron by a minor x-ray line with a slightly different energy than the main line. This results in a "shadow" peak on the low binding energy side of the photoelectron lines at a fixed binding energy offset with a magnitude proportional to the main photoelectron line (see Fig. 6) (Ref 28).

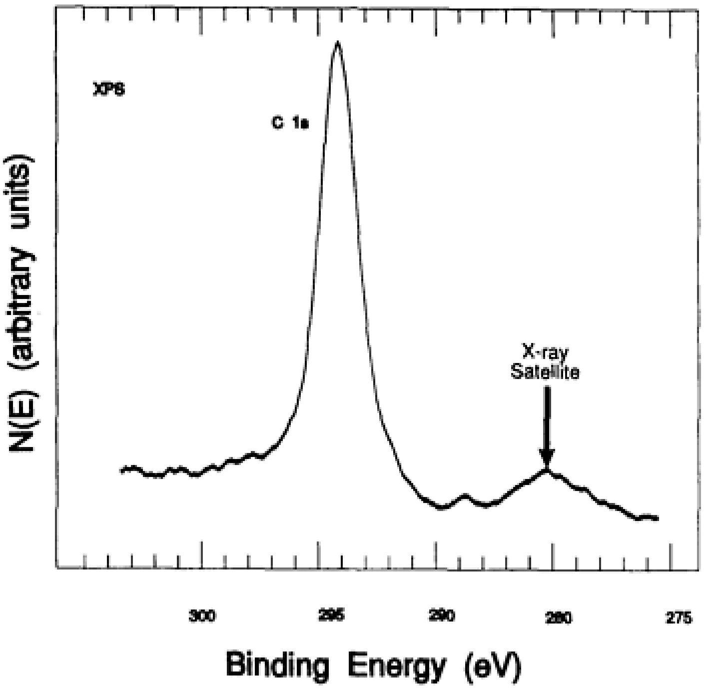


Fig. 6 Sample spectrum of C 1s photoelectron line showing x-ray satellite displaced 8.4 eV from main peak corresponding to most dominant satellite of Mg anode x-ray source. N(E) = $N_{(E)}$. Source: Ref 25

Shake-up satellites occur when the ion rear-ranges its outer electrons as the photoelectron exits in a way that leaves one or more of the valance electrons in an excited, rather than ground, state. Because this excitation energy is then not available to the photoelectron, a small peak appears at a slightly lower kinetic energy (higher binding energy). Both of these satellite peaks are usually ignored, although the shake-up peak is sometimes useful in chemical state determination when the shift in the core line is undetectable (Ref 29).

XPS Data Analysis. The first step in analyzing XPS data is to identify all species present. This begins with the calibration of the energy scale. The energy balance needs to be corrected for the spectrometer work function and sample charging. This is best accomplished by choosing some species in the sample as a reference. Adventitious carbon is often chosen, and its 1s peak is assigned a binding energy of 284.6 eV. However, this may not be a good choice, because the form of the carbon can vary widely from sample to sample. The inherent uncertainties in the spectrometer energy calibration generally limit the accuracy of the energy scale to about 0.1 eV, but this is adequate for chemical state identification of most species. Detailed descriptions of the energy calibration procedure are given in Ref 20.

After the energy scale is properly calibrated, all of the detectable species are identified by systematic identification of the peaks in the spectrum. This is usually best accomplished by first identifying the most prominent peak using reference spectra (Ref 21). Once the element from which the photoelectron line came is identified, all of the other lines, including Auger lines that are associated with that particular element, are located and identified. If any of the peaks (from that element) that should lie above the detection limit are not found in the spectrum, then the identification of the original peak is incorrect and another element is chosen. This process continues until all of the peaks in the spectrum have been identified.

Next, the identification of chemical state can be accomplished from more detailed spectra. The peak location is compared with reference data (Ref 22) and data are obtained from pure compounds. An example of the power of XPS to determine chemical states was shown previously in Fig. 2.

The quantitative analysis of XPS spectra is based on the theoretical description of the intensity (area under the main peak, plus the area under shake-up peaks, if present) of the photoelectron line, which can be accurately approximated by (Ref 5, 20, 21):

$$I_i = S_i N_i \qquad \textbf{(Eq 2)}$$

where I_i is the intensity of the peak of interest for species i, S_i is the sensitivity factor for peak i, and N_i is the concentration of species i in the sampled volume.

Sensitivity factors are tabulated for CMAs and hemispherical analyzers (Ref 21), but because the sensitivity factor is a function of the type of spectrometer, these tables may not be accurate for different types of spectrometers. The composition of the sampled region can be obtained by normalizing I_i/N_i for the major peak of each element present to the sum of the same quantity for all elements (Ref 7, 21). The accuracy of this method is typically on the order of 10%. Therefore, XPS is a semiquantitative technique with respect to composition determination. Matrix effects are primary factors that affect composition determination by using relative sensitivity factors. This is discussed more in the Auger Electron Spectroscopy (AES) section. As previously mentioned, the practical detection limit for quantification is 0.1 to 1 at %.

XPS spectra can yield structures that are composites of at least two individual peaks. This can can be the result of the presence of at least two elements in the sample that yield peaks with binding energies that are very close together. An example of this is excitation of Ta 4p and N 1s level electrons with a magnesium anode. The respective binding energies of these peaks are 403 and 400 eV (Ref 21). The details of the low-energy end of the Ta 4p structure and high-energy end of the N 1s structure can become obscured. A similar phenomenon can be observed when mixed oxidation states of a metal are present.

If, for example, the XPS spectrum of a metal consists of a doublet (two peaks), then the peaks are typically shifted to slightly higher binding energies as the degree of oxidation increases. If a sample contains the metallic species in addition to the metal oxide, then the two peaks associated with the oxide are typically shifted to slightly higher binding energies with respect to the metal. As a result, it is possible for the lower binding energy peak of the oxide to be very close to the higher binding energy of the metal. The result is a structure that is composed of two peaks.

Composite peak structures resulting from peak overlap are dealt with through either deconvolution or curve fitting techniques; the latter is probably the more popular of the two. It is important to note that neither technique can define a unique combination of single peaks that results in a given composite structure. Therefore, neither deconvolution nor curve fitting can yield a unique solution for the combination of individual peaks that results in a composite structure. A number of textbooks on spectroscopy, including those XPS resources cited in this article, describe deconvolution and curve fitting. For the purpose of this discussion, it is sufficient to point out that modern surface analysis software typically features curve-fitting routines to address the problem of peak overlap.

Ion-Scattering Spectrometry

Ion-scattering spectrometry (ISS) is based on energy analysis of ions reflected from a solid target material. The scattered species can be either inert or reactive in the most general sense, but inert species, such as argon and helium, are typically used.

The application of ISS as a surface-sensitive analytical technique is based on monitoring the kinetic energy of reflected probe ions for which the mass, incident energy, and angle of reflection are known. From this, the struck target atom mass can be determined and surface constituents identified. ISS can be applied to both electrically conductive and insulating materials, but can induce surface damage. Surface coverages near 0.01 monolayer are often detectable with ISS. At this point, ISS has not been employed in the field of friction, lubrication, and wear to the extent that XPS and AES techniques have.

ISS Fundamentals. Various articles that describe ISS fundamentals and applications have been published since 1970. Publications by Smith, Suurmeijer and Boers, and Taglauer and Heiland are excellent reference works (Ref 30, 31, 32).

In practice, the investigator is interested in using ISS to identify atomic species in the surface and near-surface regions. This is done by applying the laws of conservation of energy and momentum to the classical two-body collision model

(Fig. 7). Several points are pertinent for understanding the two-body collision process as applied to ISS. First, an incoming probe ion of mass, velocity, and energy values of M_1, V_0, and E_0, respectively, collides elastically with a target atom of mass, M_2. Second, following the elastic collision, the probe ion is deflected at some scattering angle, θ, with respective mass, velocity, and energy values of M_1, V_1, and E_1. The target atom recoils at an angle, ψ, with respective mass, velocity, and energy values of M_2, V_2, and E_2.

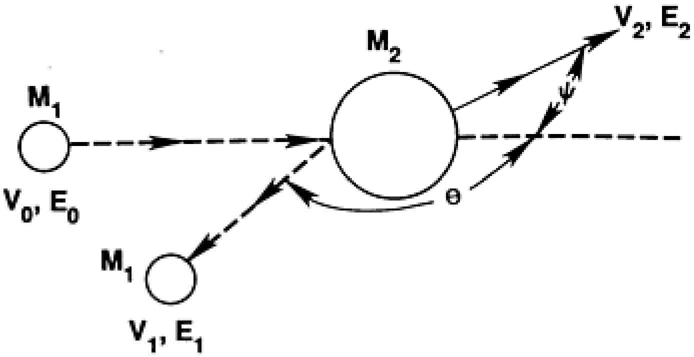


Fig. 7 Representation of two-body elastic collision between incoming atom of mass, M_1, velocity, V_0, and energy, E_0, with a target atom of mass, M_2, of initial velocity and energy of zero (stationary atom). θ and ψ represent the scattering and recoil angles, respectively.

Several equation forms are used to represent the two-body collision process. Such equations are the result of applying the laws of conservation of energy and momentum to the collision process. A typical form is that suggested by Suurmeijer and Boers (Ref 31). Specifically,

$$E_1/E_0 = \frac{1}{(1+\mu)^2}\left[\cos(\theta) + (\mu^2 - \sin^2\theta)^{1/2}\right]^2 \quad \textbf{(Eq 3)}$$

where $\mu = M_2/M_1 \geq 1$.

This form of the equation is for back scattering (that is, $M_2 \geq M_1$), as opposed to forward scattering when $M_1 > M_2$. When $M_1 > M_2$, the probe ions are not back scattered, but essentially, transfer all energy to the lighter target atoms, which recoil. It is this recoil energy that is measured in the case of forward scattering.

In practice, the forward recoil measurement is achieved by assuring a very shallow angle of approach between the target and incident ion beam and arranging the detector in a forward position in a manner such that the signal is being collected at a very shallow angle. This technique allows for detection of very light elements, with reasonable resolution when heavy probe ions are used.

Analytically, the forward-scattering process can be described by:

$$E_2/E_0 = \left(\frac{\mu}{(1+\mu)^2}\right)\cos(\psi) \quad \textbf{(Eq 4)}$$

where ψ is the recoil angle of the target mass, M_2, and $\mu = M_2/M_1$.

The scattering of ions from a solid-state surface is often classified according to the energy of the primary (probe) ion beam. Classifications of low-, medium-, and high-energy ion scattering, or reflection, are typically used. The purpose here is only to be concerned with low-energy ion-scattering spectrometry (LEISS), because of its high surface sensitivity. The

low-energy range is approximately 100 to 10,000 eV, whereas the upper end of the high-energy range is typically considered to be on order of 3 MeV.

LEISS is truly a surface analytical technique in that ions are reflected within approximately one monolayer. Probe ions with energies greater than the LEISS energy range will penetrate the sample more deeply. The result is a back-scatter spectrum with less surface-specific information.

It is important to point out that higher-energy studies can be useful for applications of thin-film analysis and analysis of subsurface regions. In order to accomplish this with lower-energy ISS, ion sputtering must be employed for "layer-by-layer" removal of the target material (see the discussion on ion etching in the Auger Electron Spectroscopy (AES) section of this article).

ISS Equipment. The equipment required to perform LEISS includes an ion source, an energy analyzer, and a vacuum system. Because these requirements closely parallel those of related surface analytical techniques, such as XPS and AES, it is not surprising that these techniques are often housed within a common vacuum system.

In some cases, the ion gun used for sputtering, or ion etching, with AES and XPS can be used as the ion source for LEISS. Historically, relatively simple electrostatic energy analyzers have been employed for ISS (Ref 30). The commercially available cylindrical mirror analyzers and concentric hemispherical analyzers typically used for AES and XPS analyses, respectively, are now popular analyzers for ISS applications.

Neutralization of Ions. Not all primary ions that are reflected from a solid target material are detected in ion-scattering spectrometry, because many ions are neutralized before exiting the target material. The fact that the neutralization probability is high for ions originating beyond approximately the first monolayer is the reason for the high surface sensitivity of LEISS. Ions are observed directly, but neutral species cannot be detected unless they are first ionized. Unfortunately, it is difficult to quantify ion neutralization processes; therefore, it is difficult to obtain quantitative information on chemical composition with ISS.

The ISS spectrum amounts to a plot of signal intensity versus E_1/E_0 (E_0 is the incident ion energy and E_1 is the reflected ion energy). If a relatively low-energy primary ion (that is, from 500 to 1000 eV) is used to bombard a sample, a relatively sharp back-scattered peak, such as shown by curve A of Fig. 8, typically results. However, as the primary beam energy is increased, a corresponding amount of skew is typically observed on the low-energy side of curve A. This is shown by curves B and C, which represent curves for respectively higher primary beam energies.

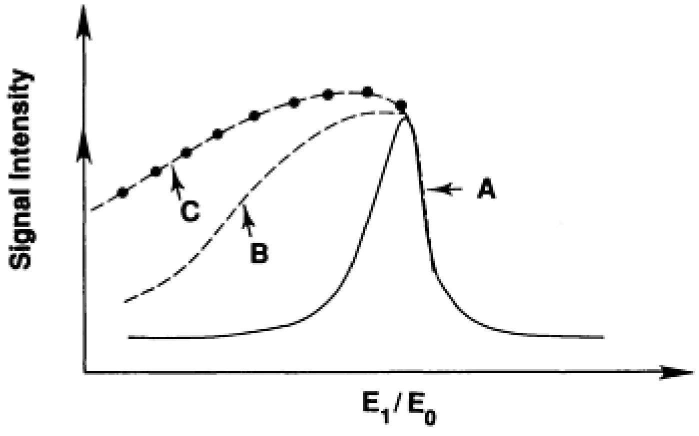


Fig. 8 Effects of deviation from single-event elastic collision model on ISS peak shape

The behavior of curves B and C, relative to A, can be explained as follows. The shape of curve A represents a low-energy beam that is reflected in a classical two-body fashion and is either elastic or very near elastic in nature. In other words, the low-energy beam undergoes true "single" scattering events. On the other hand, the higher-energy beam penetrates the sample more deeply before beginning the return path to the surface. As a result, "multiple" scattering events occur, and the energy loss of emitted ions finally escaping the surface is characterized by the broad low-energy components of

curves B and C. In fact, it is possible to observed distinct peaks at the low-energy end of the spectrum that are the result of species sputtered off of the surface, as opposed to ions that have been reflected.

The effects of multiple collisions can greatly affect the form of the ISS spectrum obtained. Interpretation of such results can be very involved; Ref 33, 34, 35 extensively discuss this topic.

In general, it is advisable for the investigator to employ the lightest possible primary ion and the lowest energies that are practical. As a rule of thumb, this means using He^+ as the ion probe, at energies up to 1000 eV. Again, results differ for the different systems and the literature should be consulted.

Resolution and Sensitivity. it is reasonable to contemplate the effects of varying the mass of the probe ion on the ISS spectrum. This is done analytically by differentiating E_1/E_0 of Eq 3 with respect to the target mass, M_2. To obtain higher mass resolution, the value of $d(E_1/E_0)/dM_2$ must be increased. Physically, this is achieved by choosing heavier mass values of M_1 for the scattering situation of $M_1 < M_2$. In fact, this idea has been employed to distinguish isotopes such as ^{63}Cu and ^{65}Cu (Ref 36). An excellent example increased mass resolution is the work presented by Smith (Ref 30), where a gold-nickel alloy with molybdenum contamination was studied. ISS was performed at E_0 = 1.8 keV for the primary ions of He^+ and Ne^+. The heavier Ne^+ primary ion was observed to result in increased peak resolution, compared with the lighter He^+ primary ion.

The sensitivity of the LEISS technique depends to some extent on the conditions of analysis. The relative intensity of a LEISS peak corresponding to an arbitrary atom, A, relative to another arbitrary atom, B, is given by:

$$\frac{I_A}{I_B} = \left(\frac{d\sigma_A}{d\sigma_B}\right)\left(\frac{C_A}{C_B}\right) \quad \text{(Eq 5)}$$

where I_i is the signal intensity of element *i*, $d\sigma_i$ is the differential cross section of element $_i$, and C_i is the surface concentration of element *i.*

The intensity, in turn, depends on the angles of incidence and scattering, and neutralization probability, as well as several other factors. Hence, by the geometry of the scattering experiment alone. It is possible to vary sensitivity, which can be as low as 0.01 monolayer (Ref 37). Spatial resolution is limited by the size of the ion beam, which is typically on the order of 1 mm (0.04 in.) in diameter.

ISS Applications. One application of ISS is well demonstrated by the behavior of polycrystalline platinum on a titania substrate. This example is particularly interesting, because under certain conditions it has been postulated that the titania substrate effectively encapsulates the polycrystalline platinum overlying titanium (Ref 38, 39, 40).

Supported transition metal oxide systems have enjoyed a great deal of attention over the years in terms of applications in heterogeneous catalysis. In the 1970s, supported platinum catalysts were extensively investigated to understand their role in methanation reactions. For such reactions, carbon monoxide and hydrogen are adsorbed on a catalytic surface and react to form methane.

Some investigators observed that when the catalyst is heated under conditions of sufficient vacuum, hydrogen and carbon monoxide no longer adsorb on the catalytic surface (Ref 41). The reason for the loss of adsorption has been a point of considerable debate. One hypothesis is that heating under reducing conditions results in the encapsulation of the metal surface by the supporting substrate, causing physical blockage of adsorbate sites. This hypothesis is sometimes referred to as the "encapsulation theory."

This theory is one of several posed to explain the behavior of reforming catalytic materials composed of systems containing platinum and titania. Under normal conditions, hydrogen and carbon monoxide are adsorbed on the platinum surface and methane is formed from their reaction. However, if the system is heated under sufficient reducing conditions, then adsorption of hydrogen or carbon monoxide is not observed. LEISS is an excellent technique by which to investigate the phenomenon, because encapsulation effects are expected to be surface-specific in nature.

The AES data shown in Fig. 9 are of a cleaned polycrystalline platinized-titania sample. The spectrum clearly shows that platinum, oxygen, titanium, and a small amount of graphitic carbon are present. Figure 10 represents ISS data for the

same sample. Peaks corresponding to single-event elastic scattering of He^+ from the sample show that oxygen, titanium, and platinum are present.

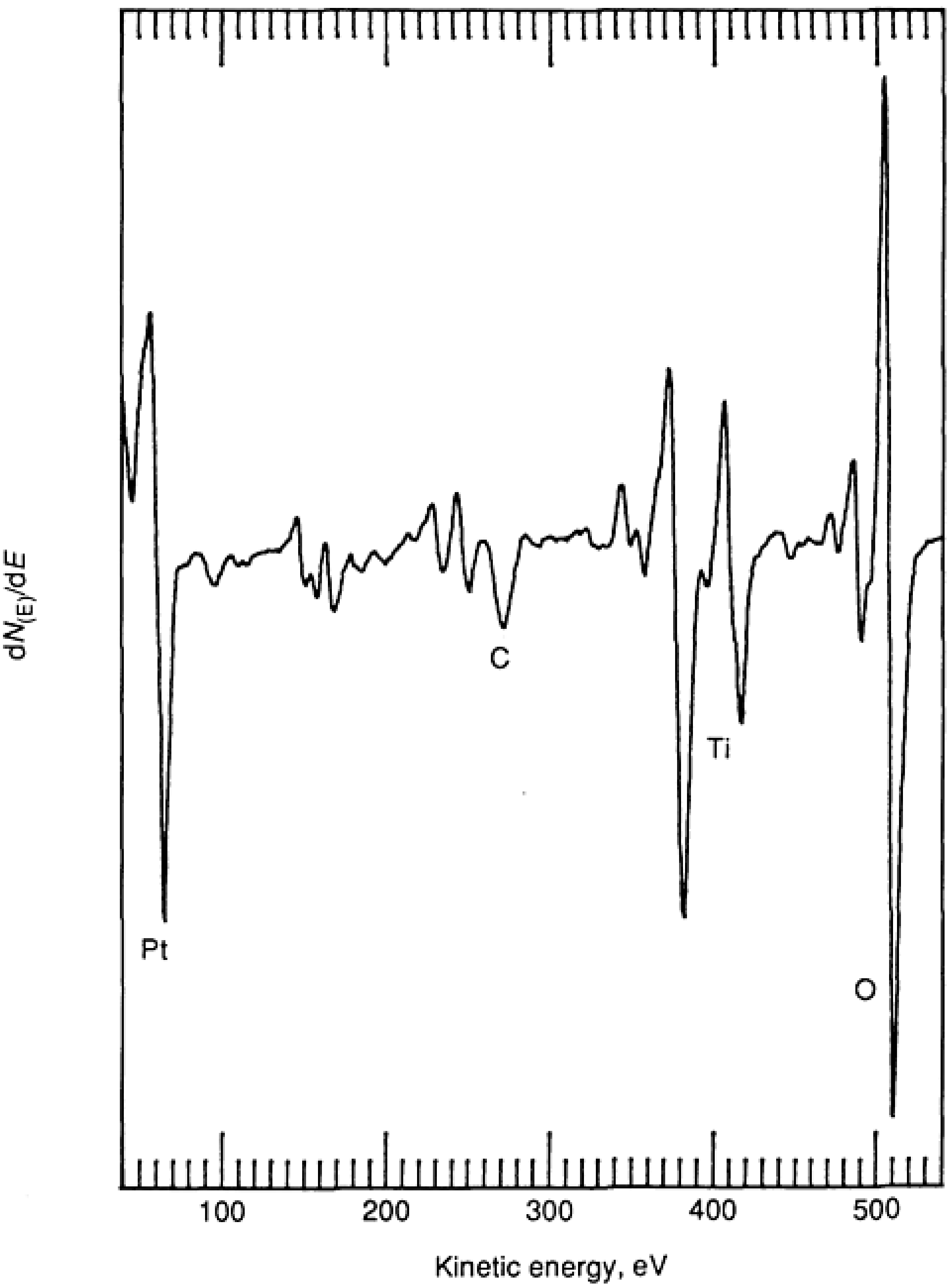


Fig. 9 AES spectrum of platinized-titania sample prior to annealing treatment

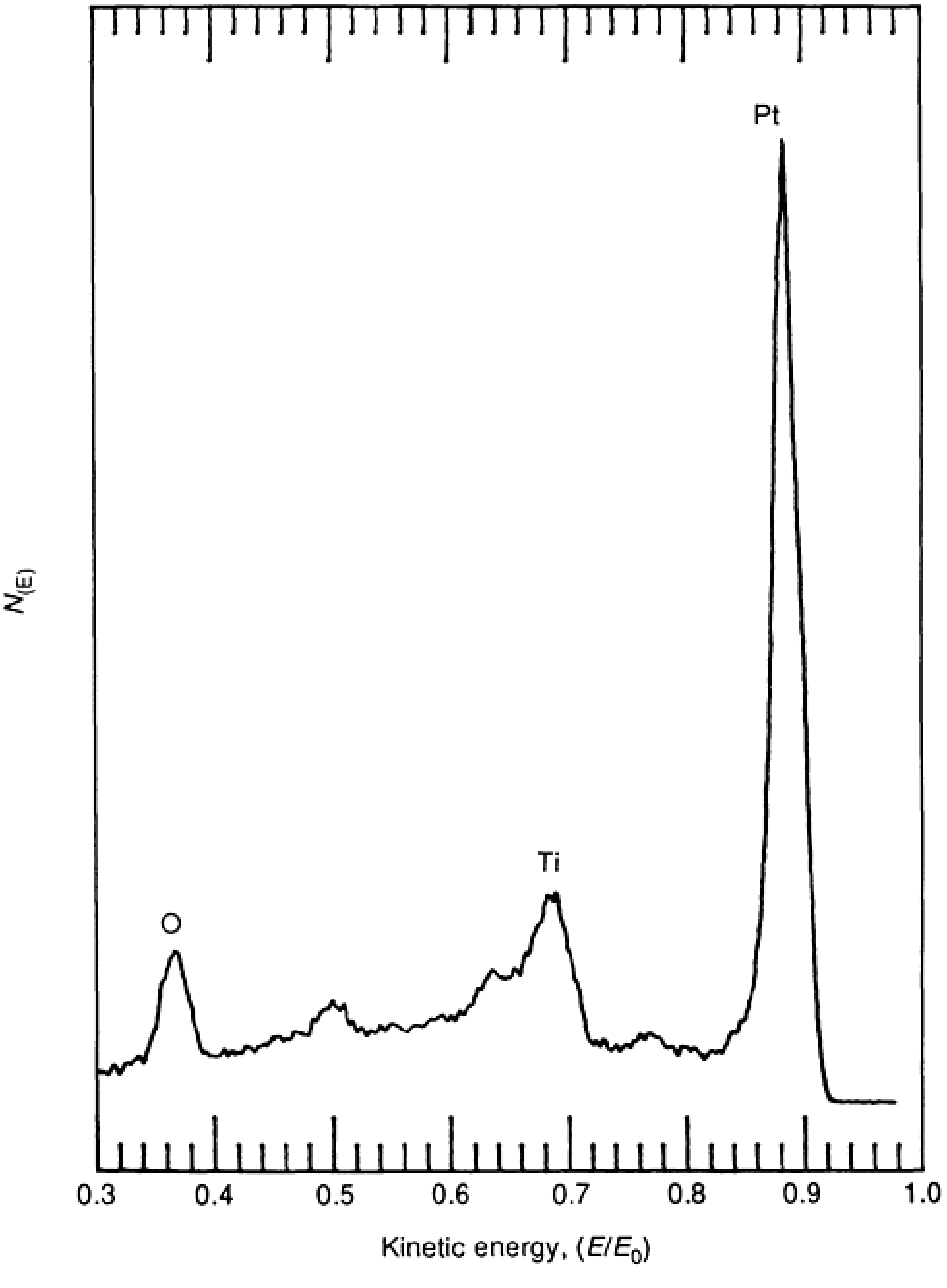


Fig. 10 ISS spectrum of platinized-titania sample prior to annealing treatment

Figures 11 and 12 are AES and ISS spectra of the same sample after annealing in ultrahigh vacuum (1.3 × 10^{-8} Pa, or 10^{-10} torr) at 500 °C (930 °F) for 1 h. Note that the relative amounts of titanium to platinum have changed, and are greater following the annealing treatment, which indicates that platinum is being overlayered by the titanium oxide substrate. Hence, ISS has been used in conjunction with AES to monitor surface and near-surface composition changes as a function of specific treatment conditions. In this case, the surface sensitivity of ISS makes the process of platinum encapsulation easy to observe.

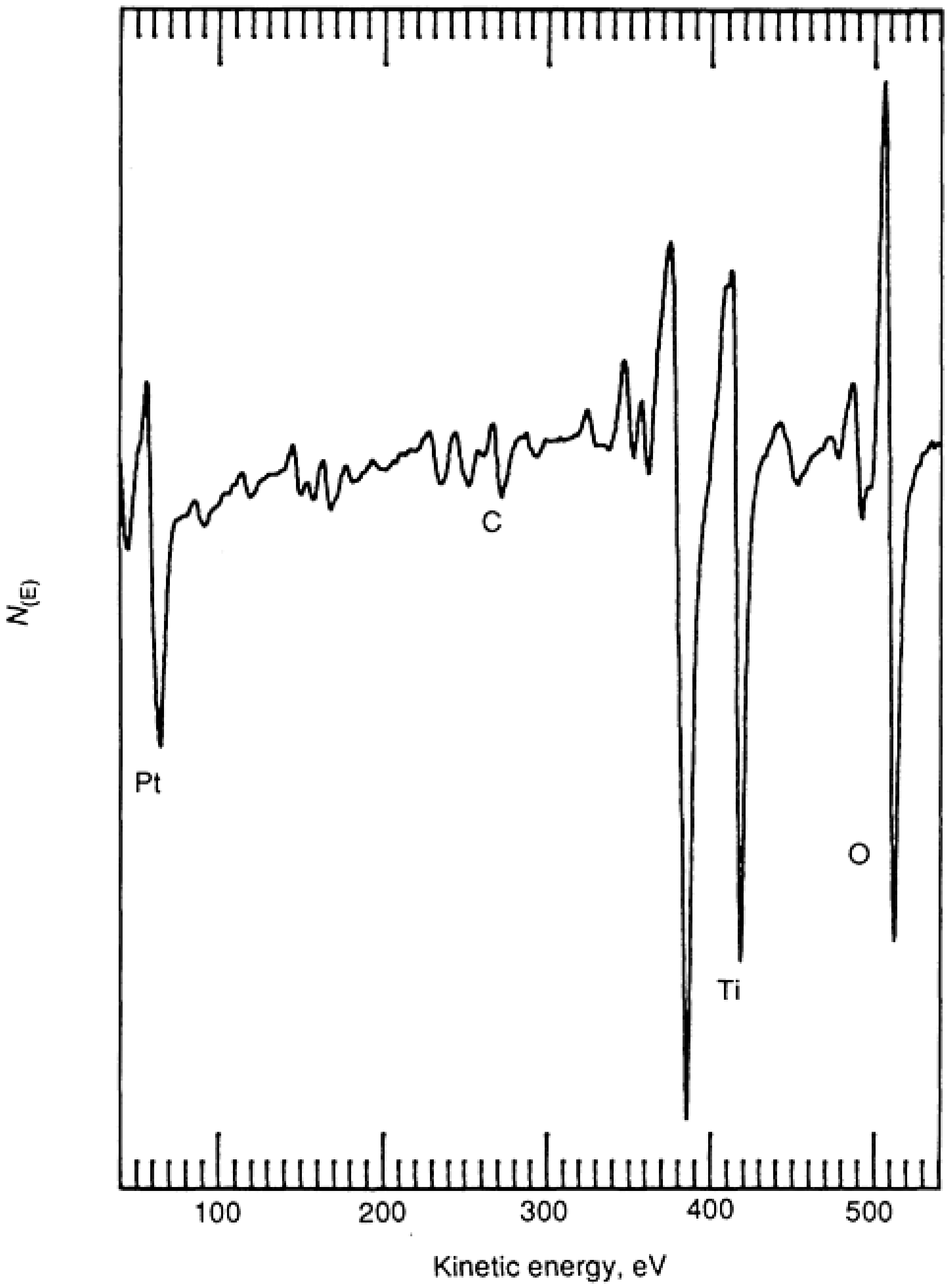


Fig. 11 AES spectrum of platinized-titania sample following annealing in vacuum for 1 h at 500 °C (930 °F)

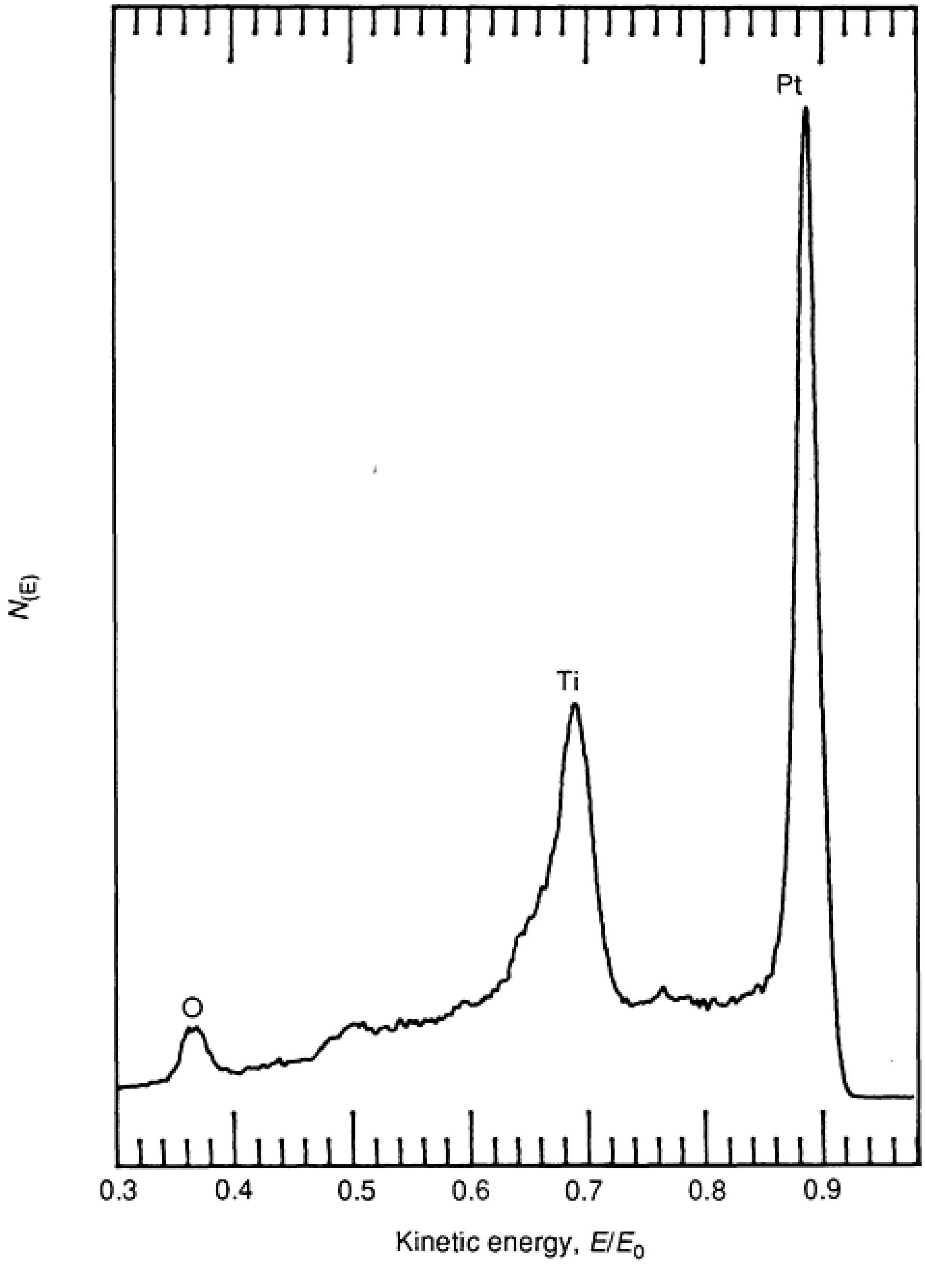


Fig. 12 ISS spectrum of platinized-titania sample following annealing in vacuum for 1 h at 500 °C (930 °F)

ISS can sometimes be applied to determine surface structure, as well as composition. For example, Niehus and Bauer (Ref 42) applied ISS to a study of oxygen adsorption on W (110). In this case, He^+ scattered at energies less than 500 eV was observed to be scattered only from the outermost atomic layer of the surface. It was therefore possible to distinguish the unreconstructed surface, consisting only of oxygen atoms, from the reconstructed surface, consisting of oxygen and tungsten atoms. Such structural studies are often done in conjunction with low-energy electron diffraction (LEED), which is a technique for structural determination. Reference 40 describes this technique.

Thus, ISS is an important analytical tool because of its high degree of surface sensitivity. In addition, it is sometimes possible to acquire structural information, especially in conjunction with techniques such as LEED. Disadvantages are the inability to obtain chemical state information and surface damage. ISS can be particularly useful in wear studies to investigate material transfer, particularly if small amounts of material are transferred and/or material transfer is restricted to very outer surface regions.

Auger Electron Spectroscopy (AES)

AES is a nonradiative spectroscopy that is based on measuring the kinetic energy of secondary electrons emitted from a solid surface when the solid is subjected to an appropriate energy source. The source is typically an electron beam with energies that range from 2 to 10 keV. However, photons and ions can also stimulate Auger electron emission from solid-state materials. The escape depth of an Auger electron depends on the energy of the electron, but is typically within a depth of five monolayers, with a detection limit of approximately 1 part per 1000. AES has probably been applied to problems in friction, lubrication, and wear more than any other surface analytical technique. Representative applications are described in Ref 43, 44, 45.

AES Fundamentals. Auger electron emission depends on the initial removal of electrons from the inner shells of an atom. This process is often referred to as core-level ionization. When a core-level vacancy is generated in an atom, the atom undergoes an electronic rearrangement, or relaxation process, through which an electron in a higher-lying energy level (lower binding energy) drops down to fill the core-level vacancy.

In the case of the Auger transition, the energy associated with this transition is imparted to a third electron is expelled from the solid, and is known as the Auger electron. The Auger $KL_{2,3}L_{2,3}$ emission process is shown in Fig. 13.

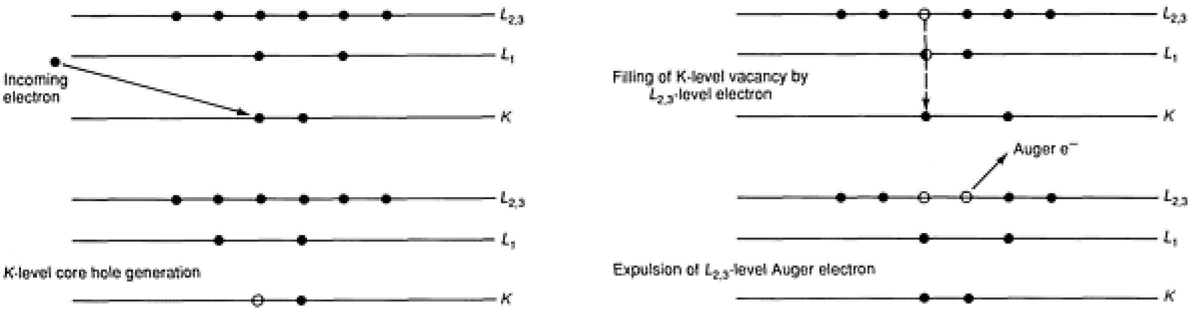


Fig. 13 Energy-level diagram at Auger $KL_{2,3}$, $L_{2,3}$ electron emission process

It is important to note that when a core-level vacancy is generated, it is also possible for the excited atom to relax via x-ray emission, as opposed to the radiationless Auger process. The emission process of an x-ray photon is the basis for electron probe microanalysis (EPMA), which is sometimes used in conjunction with AES. Unlike AES, EPMA is not a surface-sensitive technique.

Auger transitions are typically described in terms of conventional x-ray notation. A detailed discussion of this notation can be found in Ref 46 and 47. In short, x-ray notation is based on the principal quantum number, n, the orbital momentum quantum number, l, and the spin quantum number, s. The reader may recall that electrons with the same value of n occupy the same energy level or shell where n can take on the values of 1,2,3,. . .. Each principal energy level (designated by n) has one or more sublevels associated with it. The sublevels are designated by the letter, l, which is called the angular momentum quantum number. The allowed values of l are 0,1,2,3 . . . (n - 1) where n is the principal quantum number. Lastly, the spin quantum number represents the spin orientation of the electrons and this quantum number can take on a value of $s = \pm\frac{1}{2}$.

In x-ray, the principal quantum number assigned to an electron is designated by an upper case letter. In this notation system the designation of energy levels by K,L,M,. . . corresponds to values of n = 1,2,3, . . . for the principal quantum number. Next, each upper case letter is assigned a numerical suffix value, each of which represents a specific combination of angular and total momentum values. The total momentum is designated by the letter j and is the sum of the orbital and spin momentum (that is, $j = l + s$). A suffix value of l specifies $l = 0$ and $j = \frac{1}{2}$, while a value of 2 specifies $l = 1$ and $j = \frac{1}{2}$, a value of 3 specified $l = 1$ and $j = \frac{3}{2}$, and so on. For example, the x-ray notation of L_1 means $n = 2$, $l = 0$, and $j = \frac{1}{2}$, while the L_3 notation means $n = 2$, $l = 1$, and $j = \frac{3}{2}$.

Electron Escape Depth. The probability that an electron will be emitted from a solid, following generation, is related to the mean free path, λ, and the below-surface depth where the electron originates. Specifically, the probability of escape, *P*, can be written as:

$$P = \alpha \exp\left(\frac{-z}{\lambda}\right) \quad \textbf{(Eq 6)}$$

where *z* is the depth below the surface, λ is the mean free path of the electron, and α is a proportionality constant. In turn, the mean free path is a function of the energy of the electron. This path is relatively insensitive to the material through which the electron is moving and depends only on the energy of the electron. As a result, universal curves have been developed to graphically show the relationship between electron mean free path and electron kinetic energy. Such curves can be obtained from Ref 46.

As the Auger electron emission process begins, it is clear that several things can happen to the electron as it travels through the solid. The further the electron travels, the greater the probability that it will interact with other electrons in the solid. If this happens, the Auger electron will lose energy and eventually emerge at some energy less than the initial level. Clearly, only Auger electrons originating in the surface and near-surface regions eventually escape the solid and are recognized at characteristics Auger energies. These electrons that leave the solid and are then detected originate from an average distance, *Z*, below the surface, where *Z* is typically less than 10 monolayers (Ref 47). As a result, the AES spectrum peaks associated with these electrons are often referred to as elastic peaks.

It is important to point out that energy loss processes also occur as the result of interaction of emerging electrons with phonons and plasmons. Discussion of these phenomena can be found in Ref 48.

AES Equipment. Chemical analysis by AES involves the measurement of characteristic energies of electrons emitted from the surface and near-surface regions of solid-state materials. Hence, the fundamental requirements of the overall process consist of an appropriate energy source to stimulate Auger electron emission, a UHV system to maintain sufficiently low pressures, and an analyzer to measure energies of emitted Auger electrons.

The electron gun typically contains a tungsten filament that produces electrons through thermionic emission. In the case of scanning Auger instruments, beam sizes on the order of 1 μm are possible. The electron gun can theoretically be positioned in any manner that proves to be practical. However, for energy analysis, many systems employ cylindrical mirror analyzers that house the electron gun coaxially.

The vacuum system, which must be operated in the UHV range, has a stainless steel "bell jar" for the main chamber. System base pressures are typically on the order of 1.3×10^{-8} Pa (10^{-10} torr), with pressures during operation being slightly higher. Operational pressures in the range of approximately 6.5×10^{-8} to 1.3×10^{-6} Pa (5.0×10^{-10} to 1×10^{-8} torr) are typically encountered.

Three types of pump systems are usually employed to bring the system down to the desired pressure. Either mechanical roughing pumps or liquid nitrogen sorption pumps are used to initially start the pump-down process. Pressures on the order of 0.13 Pa (10^{-3} torr) are typically achieved with this primary pumping process. Next, turbomolecular pumping is typically employed to achieve pressures on the order of 1.3×10^{-3} Pa (10^{-5} torr), followed by ion pumping, which brings the system into the UHV range.

Eventually, contaminants accumulate until it is no longer possible to obtain the desired system base pressure. In this case, the system must be baked to remove adsorbed contaminants from system components, at which point they are pumped away. Procedures for baking vary from system to system and equipment manufacturers should always be consulted during the first baking procedure. Failure to execute this procedure correctly can result in equipment damage.

Energy Analyzer. The most popular type of analyzer for AES applications is probably the cylindrical mirror analyzer, previously discussed in the section on XPS. The CMA consists of two coaxially arranged cylinders. Electrons from the target material are electrostatically focused such that only electrons with specific energies will successfully traverse the length of the analyzer and contact the electron detector at the back of the analyzer.

The focusing process is accomplished by applying a voltage across the two cylinders. The outer cylinder is biased at a negative potential relative to the grounded inner cylinder. An electric field results from the voltage application across the

cylinders. The strength of the electric field is related to the magnitude of the voltage potential difference across the cylinders. As the voltage difference is increased, the field strength is increased. Only electrons within the energy spread, ΔE, have appropriate trajectories to reach the detector. Each electron has an energy, E = eV, where e is the charge on the electron and V is the applied potential. The energy analyzer resolution for this situation is defined as $\Delta E/E$.

CMAs are available in two configurations: single-or double-pass (Fig. 14). The single-pass configuration is well suited to AES because the electron beam typically used as a probe is focused over a small area relative to the entrance area of the analyzer. Therefore, the transmission of electrons through the analyzer is significant, relative to that observed for XPS, which employs a more diffuse x-ray source for excitation.

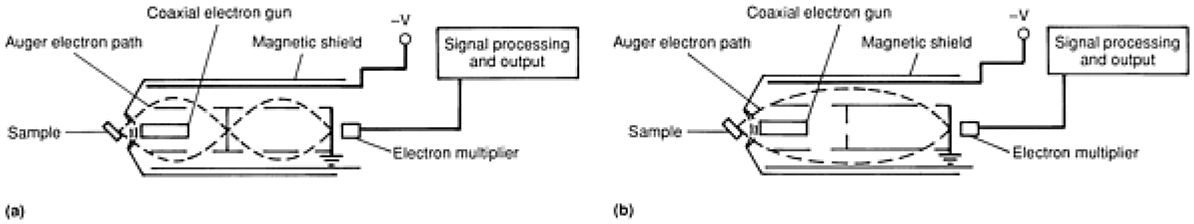


Fig. 14 Cylindrical mirror analyzer types. (a) Double pass. (b) Single pass

The double-pass CMA has found extensive application when AES and XPS are performed together. As already noted, additional focusing capability and resolution requirements are necessary for XPS, which means that this technique requires the double-pass CMA. Systems dedicated exclusively to XPS often employ analyzers of hemispherical geometry (Ref 46). The larger area over which the signal is collected with this geometry, relative to the CMA, results in detection enhancement and is therefore very attractive for XPS application.

The AES spectrum is a representation of the distribution of energies of Auger electrons emitted from a material of interest. This energy distribution is represented by displaying signal intensity, $N_{(E)}$, as a function of Auger electron kinetic energy. Therefore, AES data appear in the form of $N_{(E)}$ versus Auger electron kinetic energy, as shown in Fig. 15. This $N_{(E)}$ spectrum was obtained by subjecting a copper sample to a 5 keV electron beam. The spectrum shows that copper, oxygen, and carbon are present. Copper is characterized by three peaks in the range of 750 to 925 eV and one low-energy peak occurring near 50 eV. Oxygen and carbon are single peaks occurring at intermediate energy values. An important aspect of data in the $N_{(E)}$ form is that the signal is typically small and is contained on a large secondary electron background.

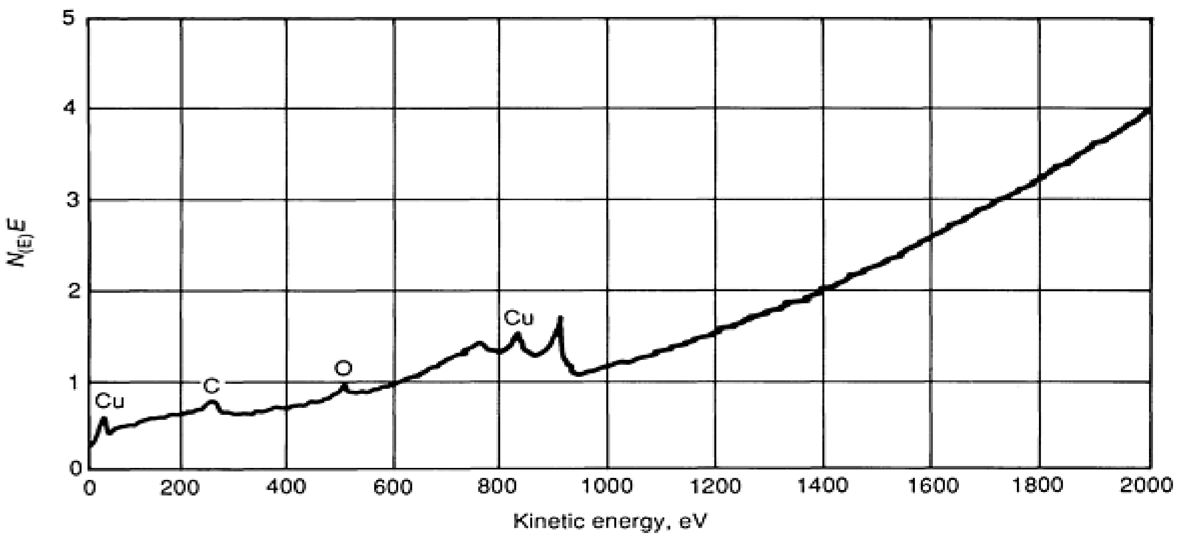


Fig. 15 AES spectrum of copper sample displayed in $N_{(E)}$ mode

The interpretation of AES data presented in the $N_{(E)}$ mode is greatly enhanced by differentiation of the $N_{(E)}$ signal. This mode of display is referred to as the d$N_{(E)}$/dE mode. The $N_{(E)}$ data in Fig. 15 is shown in d$N_{(E)}$/dE form in Fig. 16. The peak-to-peak height of signals in the d$N_{(E)}$/dE mode is proportional to the number of atoms from which Auger electrons originate to yield the signal.

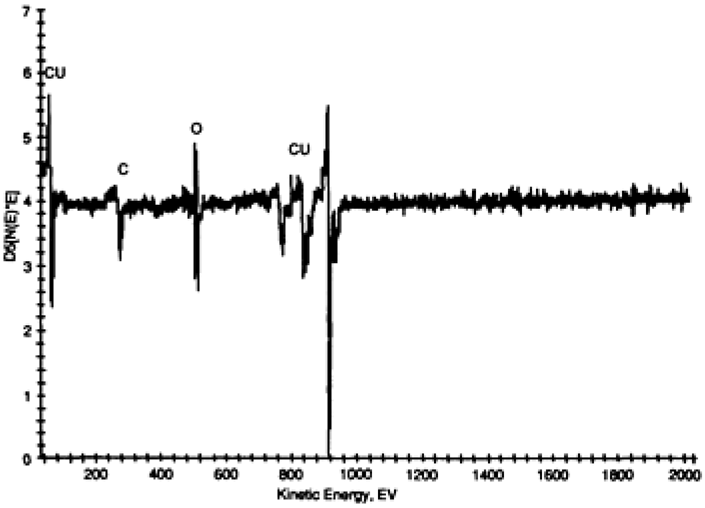


Fig. 16 AES spectrum of copper sample displayed in derivative d$N_{(E)}E$/dE mode. D5[N(E)*E] = d$N_{(E)}E$/dE

Qualitative analysis can be performed by measuring the kinetic energy of electrons over a range that encompasses emissions from all atoms of interest. In practice, the most useful range of electron kinetic energies is probably 30 to 2000 eV. Once an initial survey scan is completed, it is often useful to perform subsequent scans over more-narrow energy regions that contain peaks of interest. The smaller energy window will enhance the resolution of peaks in this region with respect to features that are not as easily observed in the more-compressed 30 to 2000 eV scan.

Quantitative results for compositional analysis are often extracted from AES data. The reader should realize that methods of quantification typically utilized in practice are best described as semiquantitative. This is because reference data are typically based on signal intensities obtained for pure reference materials. When the element is present in the more complex matrix environment of a compound, the behavior of emitted Auger electrons can be expected to be different. This is primarily manifested through complicated scattering events and changes in the inelastic mean free path. Scattering events can be significant because electrons that have encountered scattering events can sometimes contain sufficient energy to induce Auger transitions. The inelastic mean free path, which is the characteristic path length traveled by the Auger electron before it is inelastically scattered, is certainly subject to matrix effects.

As discussed in Ref 46, a linear relationship is the simplest form of relationship between the AES signal intensity of an atom of interest in pure form and in a matrix with other elements. For instance, the atomic fraction, X_i, of an atom, i, in a matrix is given by

$$X_i = \frac{I_i}{I_i^p} \qquad \textbf{(Eq 7)}$$

where I_i is the Auger signal intensity observed for the atom in the matrix and I_i^{p} is the intensity of the pure component. It should be clear that pure component intensity values for an element of interest must be either available or measured for a given set of operational conditions. In any event, a standard representing each element of interest must be measured at some point. This typically requires either preparation or acquisition of a large number of standards.

In practice, the composition of a material of interest can be determined in a couple of ways. One method of obtaining compositional information is to compare element peak intensities for a material of interest to intensities obtained from pure element standards. Alternatively, relative sensitivities of pure elements with respect to a pure element standard can

be correlated through the use of elemental sensitivity factors. In this case, the atomic fraction of species *i* in a matrix of *N* elements is given by

$$X_i = \frac{(I_i/S_i)}{\sum_{i=1}^{N} (I_i/S_i)} \qquad \textbf{(Eq 8)}$$

where X_i is the atomic fraction of species *i*, I_i is the signal intensity of species *i*, and S_i is the elemental sensitivity of species *i*.

It is important to point out that when data are collected in the derivative mode, I_i is the peak-to-peak height, and when data are collected in the $N_{(E)}$ mode, I_i is the peak area. It should also be noted that inaccuracies are inherent in the sensitivity factors because they are derived for pure elements and do not include matrix and chemical effects. Even though they are not commonly employed for outline analysis, techniques have been developed to evaluate matrix effects through the application of matrix factors (Ref 46).

Ion Etching (Sputtering). A very useful feature of most surface analysis equipment is the ability of ion etch, or sputter, the surface of a sample of interest. This process consists of eroding away the surface in a layer-by-layer fashion. The two primary uses of ion etching are acquisition of depth-dependent composition information (depth profile) and removal of contaminants from the surface of a material of interest.

The process of ion etching involves ion bombardment of a target surface with sufficient energy to break atomic bonds in the solid and remove chunks of the surface from the solid target. Ions are generated for this process by electron-atom collisions in a pressurized gas. It is usually desirable that the gas be chemically inert. Therefore, ions of the noble gases are typically employed, with Ar^+ being used most often. Once the Ar^+ beam is generated, it is delivered to the target with energies that typically range from 2 to 5 keV, with a beam diameter on the order of a couple of millimeters. Depending on the specifics of the equipment employed, the ion beam can either be applied to one spot or rastered over an area.

AES depth profile information is usually reported in the form of atomic concentration, or atomic fraction, versus sputtering time. Absolute depth values are usually not convenient to determine and can actually be very difficult to determine, depending on the particular system. Usually, ion etch rates, or sputter rates, are reported relative to a known sputter rate. In other words, a material of known thickness is selected, the time required to sputter through this thickness is recorded, and then rates for other materials are reported with respect to this rate. In most instances, the reference material used is Ta_2O_5. Historically, this choice has been made because of the ease with which the oxide film thickness can be controlled during formation (Ref 46).

Consider the AES depth profile data shown in Fig. 17. These data were acquired by sputtering with Ar^+ at 2 keV. The data are for a BiO_x-Auglass thin-film system prepared by thermally evaporating gold on a glass substrate, followed by the evaporation of bismuth over the gold. Subsequently, the sample was subjected to a treatment to oxidize the bismuth. This system was chosen to demonstrate the depth-profiling technique with AES, because the layered structure allows the clear identification of changes in composition through the layers of the thin-film system.

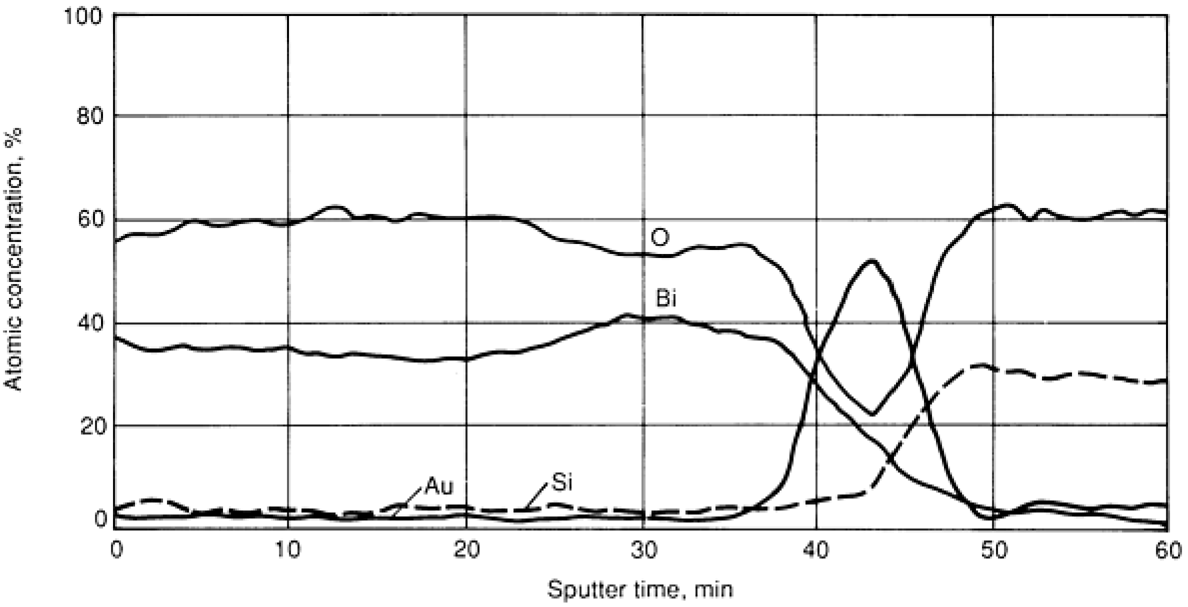


Fig. 17 AES depth profile of BiO_x-Au system on glass

It is imperative that the reader understand the complexity of the atomic processes that result in ions being etched from the surface. References 46 and 47 provide detailed descriptions of these processes. However, the potential limitations of ion etching can be summarized here. First, ion etching induces changes in the composition of the target material in local areas where etching is occurring. The degree of alteration of the true composition profile depends on the specifics of the system and conditions. Second, ion etching rates are subject to matrix, as well as grain boundary, effects. Again, the extent of these effects depends on the particular case, but can be substantial. Third, ion etching can induce surface roughness, which can affect subsequent analytical information. Last, it should always be remembered that ion etching is a destructive technique by nature and precautions should be taken if this presents a problem.

Effects of Chemical Environment on AES Spectrum. Changes in the chemical environment of an atom can affect the AES spectrum with respect to that atom. Such perturbations of the atomic environment are typically observed as shifts in the location of AES elastic peaks and/or changes in the prestructure of these peaks.

Shifts in the measured kinetic energy of AES electrons result when electron transfer occurs between atoms. The greater the extent of electron transfer, the greater this effect is expected to be. Therefore, chemical effects for systems exhibiting ionic bonding character should be important. In fact, chemical shifts of several electron volts can be observed for this particular situation. The effects of chemical shifts show in the AES spectrum as peak shifts to higher kinetic energies for electronegative atoms and as shifts to lower binding energies for electropositive atoms.

A classic case of chemical effects in the AES spectrum can be seen by comparing spectra for aluminum and alumina. In this case, both AES elastic peak locations and prestructures are significantly different. Reference 49 compares these spectra.

AES Applications. AES and related surface analytical techniques have enjoyed widespread application in many areas, although applications have been more extensive in areas such as catalysis than in areas such as friction, lubrication, and wear. However, as suggested earlier, AES has probably been applied to friction, lubrication, and wear problems more than any other method of surface analysis.

From the discussion thus far, it is clear that AES can be used to identify chemical constituents and to determine their relative abundance in terms of atomic fraction. In some cases, qualitative information about oxidation states can be obtained by AES peak structure. Changes in the shape of an AES peak can indicate a change in oxidation state. For example, a comparison between Fig. 9 and 11 shows that the AES Ti 387 eV peak shape is changed following annealing in vacuum (reducing conditions). Therefore, titanium oxide reduction can be qualitatively identified by the shape of the Ti 387 eV peak. This phenomenon has been discussed in detail by Hoflund *et al.* (Ref 50).

Consider the scanning electron micrograph shown in Fig. 18. The micrograph shows a wear track area on an aluminum 1100 surface that has been worn against an annular silicon carbide counterface in a distilled-water environment. The wear test was performed with a new test configuration that has recently been reported (Ref 51). Representative areas, both within and external to the wear track, will be analyzed as an example of the application of AES.

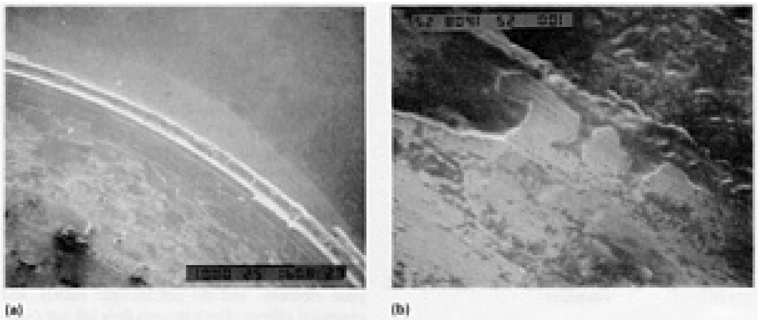


Fig. 18 Electron micrographs. (a) Wear track edge of worn aluminum 1100 sample. (b) High-magnification shot inside wear track

Figure 19 is an AES survey spectrum acquired using an Auger electron spectrometer with secondary electron imaging capabilities. The survey scan was acquired within the wear track region with electron beam voltage and current values of 5 keV and 1.67 μA, respectively. In fact, these parameter values were maintained for all data presented in this section.

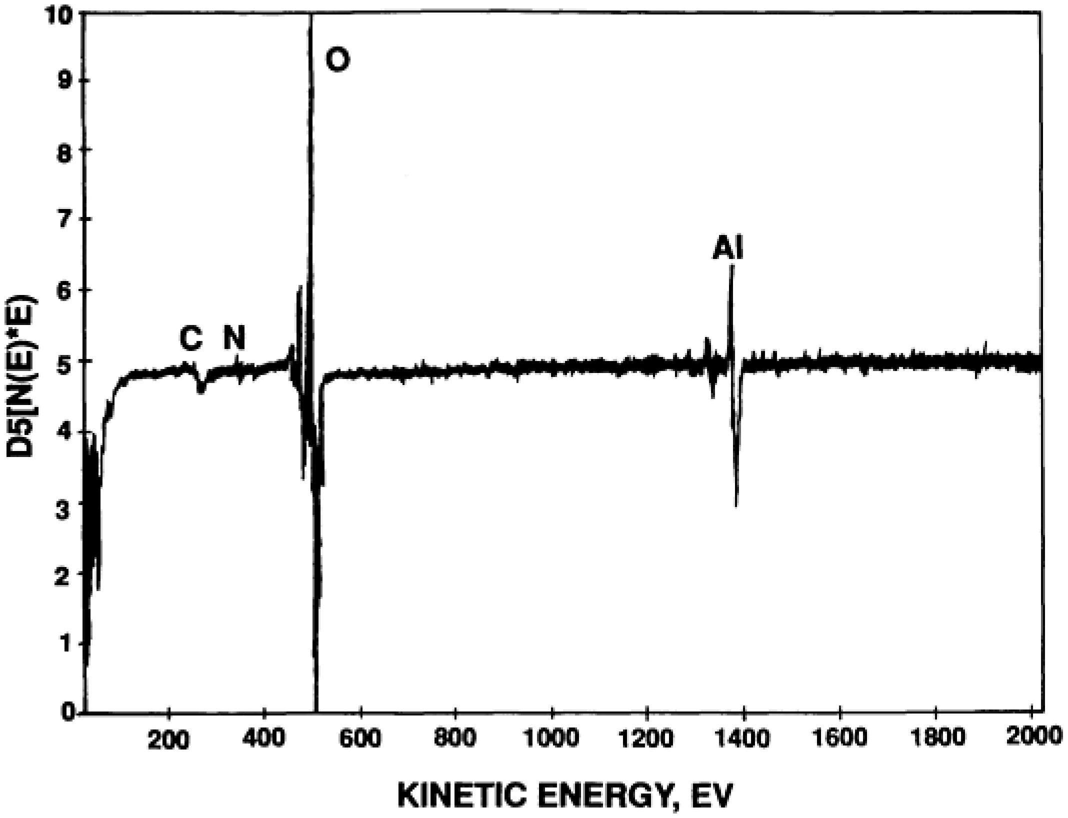


Fig. 19 AES survey spectrum of wear track region of aluminum 1100 sample. D5[N(E)*E] = d$N_{(E)}E$/dE

The data in Fig. 19 show that the surface and near-surface regions of the worn area typically contain small amounts of carbon and nitrogen, in addition to aluminum and oxygen. The carbon peak is expected to occur at 242 eV, oxygen at 503

eV, aluminum at 1396 and 68 eV, and nitrogen at 376 eV (Ref 43). The atomic fractions of aluminum, oxygen, and carbon are 0.62, 0.31, and 0.07, respectively. The small amount of nitrogen has been neglected in the calculation. The "oxide character" of metallic systems is often of interest. For this reason, the atomic ratio of oxygen to aluminum (O/Al) is noted to be 0.50.

In order to compare the composition of the specimen in regions below the wear track to the above-discussed composition in the surface and near-surface regions, ion etching was performed to effect material removal. Figure 20 shows the result of an AES survey scan acquired following ion etching for 10 min with 4 keV Ar^+. The AES spectrum shows that carbon and nitrogen have been completely removed. Only oxygen and titanium occur in this region. The small amount of argon shown by the AES signal at 215 eV is the result of Ar^+ being embedded in the specimen during ion etching. The (O/Al) value is 1.3, which is considerably more than the value of 0.50 obtained at the surface. The respective atomic fractions of aluminum and oxygen are 0.44 and 0.56. An analogous set of data was obtained for areas external to the wear track to allow a comparison between the worn area and a typical area of the specimen not subjected to the wear process.

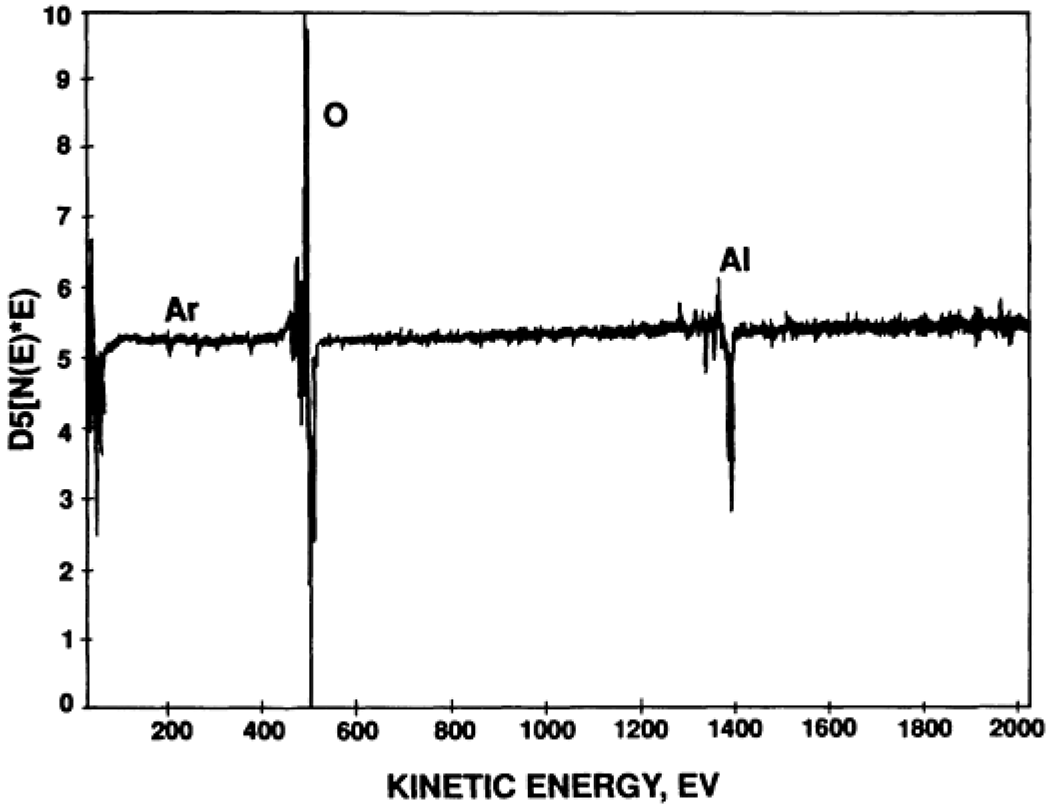


Fig. 20 AES survey of aluminum 1100 wear track region following sputtering with Ar^+ for 10 min at 4 keV. D5[N(E)*E] = d$N_{(E)}E$/dE

Figure 21 shows that chlorine, carbon, and a small amount of nitrogen, in addition to oxygen and aluminum, are present in the surface and near-surface regions. Note that the amount of carbon is considerably more than was observed in the wear track area. It is difficult to identify a single carbon source in this case. Carbon is a common component of most surfaces before the sample is cleaned by light ion etching and/or heating. Chlorine was present on the surface of some of the as-received material and therefore shows up here for the region external to the wear track area. This suggests that chlorine, nitrogen, and a considerable amount of carbon are removed during the wear process in the deionized water environment. The respective atomic fractions of aluminum, oxygen, carbon, and chlorine are 0.42, 0.24, 0.33, and 0.01. The (O/Al) value is 0.57.

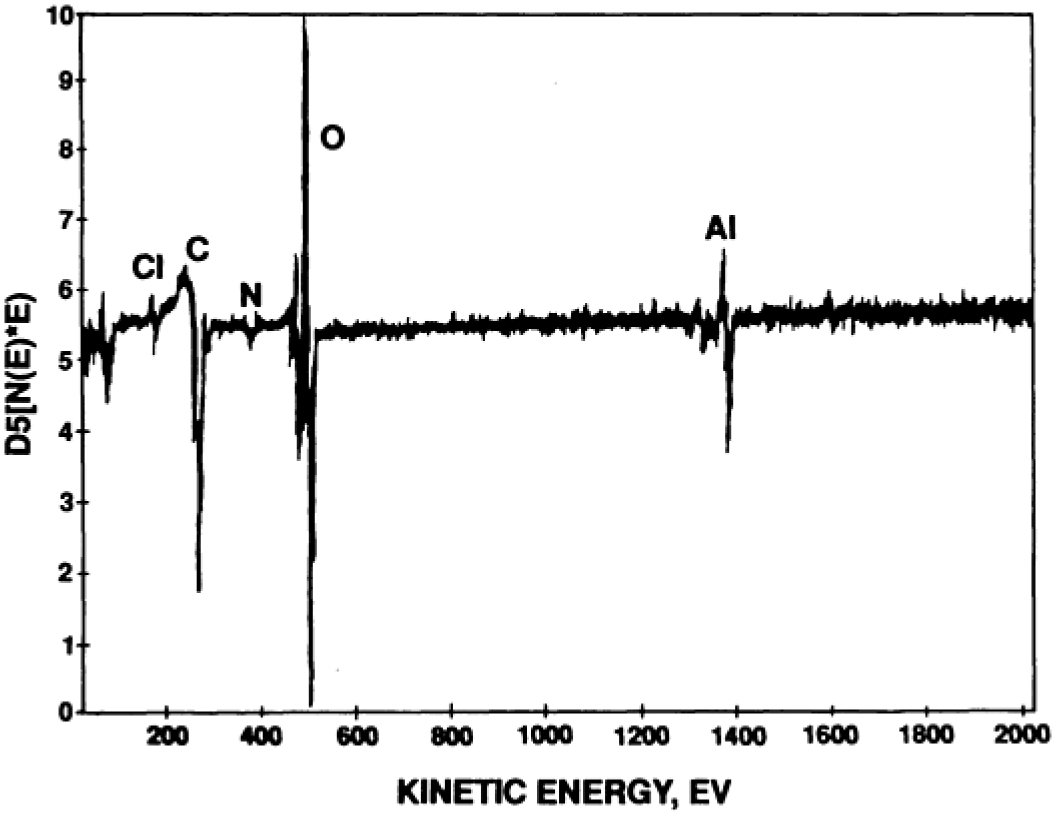


Fig. 21 AES survey spectrum of region external to wear track. D5[N(E)*E] = d$N_{(E)}E$/dE

Figure 22 demonstrates ion etching for 10 min with 4 keV Ar^+. The spectrum shows that chlorine and carbon have been effectively removed. As previously observed, argon appears in the spectrum after being embedded during the ion etching process. The (O/Al) value is 0.52, which is close to the value of 0.57 obtained prior to ion etching.

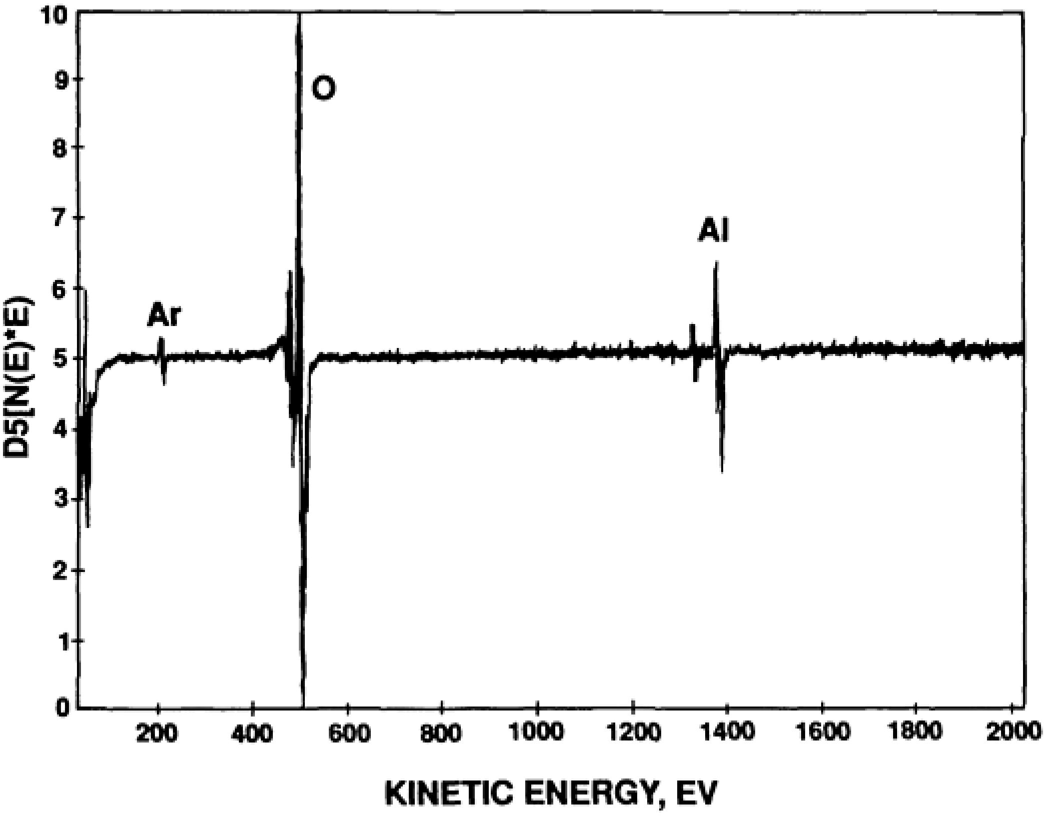


Fig. 22 AES survey spectrum of region external to wear track following sputtering with Ar^+ for 10 min at 4 keV. D5[N(E)*E] = $dN_{(E)}E/dE$

This example represents typical results obtained following evaluation of a number of sample areas, both inside and external to the wear track area. Probably the most interesting observation is that the (O/Al) value is consistent in moving from the surface and near-surface regions toward bulk regions, as deduced from the analysis of the specimen in regions external to the wear track. These (O/Al) values of 0.57 and 0.52 are also in the neighborhood of the 0.50 value observed for the surface and near-surface regions of the wear track. On the other hand, the (O/Al) value of 1.3 observed following ion etching in the wear track region is considerably higher. Even though a mechanism has not been proposed or tested here, the AES data indicate that the wear process could result in oxygen transport to interior regions of the specimen, resulting in increased (O/Al) values.

Finally, a point of caution should be observed when comparing composition data. The accuracy of data is probably at best on the order of 10%. Therefore, if it is determined that the atomic fraction of an element in a material of interest is 0.50, it should be viewed as 0.50 ± 0.10. This means that the composition should be interpreted as ranging from 0.45 to 0.55. This uncertainly stems from uncertainties in relative sensitivity factors that arise from complications such as matrix effects.

AES limitations are described below in terms of electron beam, nonconductive samples, and the overlapping of peaks.

The electron beam, typically used as the probe in AES, has several drawbacks. For instance, localized damage can occur through electron-stimulated desorption processes. As a result, composition can be altered in the subject area.

Nonconductive samples may "charge up" under the electron beam and yield spurious results. Major characteristics of surface charging are peak shifting and distortion. If the instrument has secondary imaging capabilities (scanning Auger), points of charging can sometimes be identified by their brightness.

Several procedures are commonly used to deal with the analysis of materials that demonstrate poor electrical conductivity. Higher beam voltages can sometimes be employed to alleviate the problem. Also, if shallow contact (grazing) angles are established between the electron beam and the solid, then surface charging can be minimized because this arrangement maximizes secondary electron emission from the surface. It is important to note that conventional

conductive coating techniques such as those employed with scanning electron microscopy (SEM) are not practical with AES because of the surface sensitivity of AES measurements. Also, it is sometimes possible to raster the electron beam to the extent that localized areas of charging are avoided or at least minimized.

Peak overlap is usually not a major problem with AES, but it can occur and should be addressed. This problem can be a combination of differences in relative sensitivities and concentrations of the elements involved. For instance, if the characteristic peak(s) of one constituent is very near the peak location of another constituent, and the signal intensity of one is considerably less than the other, then the constituent with the higher sensitivity can mask the peak of the other. This is especially true if chemical elements with low sensitivities are present in small concentrations.

Consider the peaks of chromium that occur at 529 eV and 489 eV in the ($dN_{(E)}/dE$) mode and the primary peak of oxygen that occurs at 503 eV in the ($dN_{(E)}/dE$) mode (Ref 49). These peak locations are such that overlap and masking can occur in some situations. For example, consider the Auger spectrum shown in Fig. 23. The spectrum is of a chromium catalyst material composed of mixed oxides of chromium and tin. Note that oxygen dominates the spectrum in the region where oxygen and chromium occur. The amount of chromium is so small, relative to oxygen, that the primary chromium peaks cannot be distinguished. In this case, one alternative would be to utilize other methods, such as XPS, to examine the chromium. It is convenient to use an x-ray source to stimulate Auger transitions when peak overlapping occurs. The resulting transitions will typically be easily distinguished (see Fig. 5).

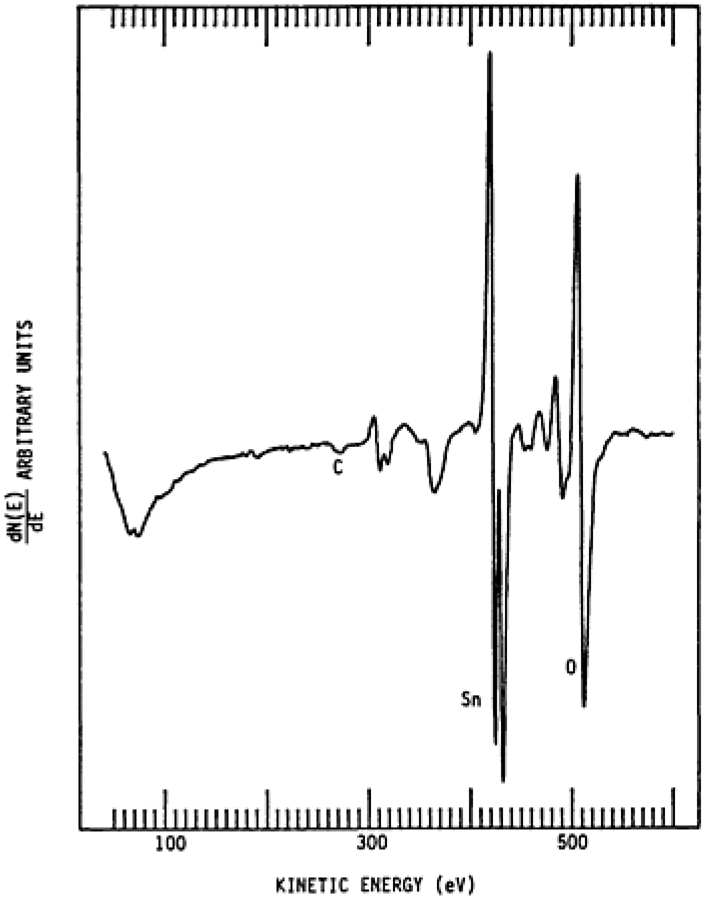


Fig. 23 AES spectrum of cleaned polycrystalline tin oxide sample

Electron-Stimulated Desorption (ESD)

Electron-induced desorption of ions is the basis of ESD. The technique, which has only recently been applied to chemical analysis, offers the advantage of hydrogen detection with minimum surface damage, compared with secondary ion mass spectroscopy (SIMS), which is an alternative technique that is usually used for hydrogen detection. A major disadvantage of ESD is that the data are generally difficult to interpret beyond the simple identification of low molecular weight ions.

The detection limit of ESD varies considerably, depending on the particular situation. In general, the detection limit of ESD is greater than that of XPS and AES. Some adsorbates, such as fluorine, can be detected when present in concentrations near 1 ppm. ESD has not been specifically applied to friction, lubrication, and wear problems at this time, to the knowledge of the authors.

ESD Fundamentals. The desorption of an ion by electron stimulation involves electronic transitions and is not the result of thermally stimulated desorption, or momentum transfer from electrons in the incident electron beam to surface atoms of the solid. Many mechanisms have been proposed for ESD, and Ref 52 provides an excellent review of this material.

Two mechanisms, in particular, have received considerable attention: the Menzel-Gomer-Redhead (MGR) model (Ref 53) and the Knotek-Feibelman (KF) model (Ref 54). The MGR model is based on electron ejection from a bonding orbital, leading to an antibonding state. This is an unstable configuration relative to the bonding state, prior to electron ejection. An excited adsorbate state can result from the antibonding state.

The MGR model can be depicted as shown in Fig. 24. Curve G represents a potential energy curve for a surface atom, such as an adsorbate atom, vibrating in the ground state in proximity to the solid surface. For the sake of discussion, the solid is assumed to be a metal and is designated by (M). The adsorbed atom is designated by (A). Curve (M + A)* represents an antibonding state of (A) following excitation by electron loss from a bonding orbital. Curve M + A* represents an excited state of (A), and the $M^- + A^+$ curve represents an ionized state where the positively charged adsorbate can desorb if the appropriate energy configuration exists. The arrow in Fig. 24 indicates how a transition from the ground state to the ionized state results in desorption of an ion. Note that the arrow intersects the potential energy curve of $M^- + M^+$ in the repulsion region. Therefore, this potential energy becomes the kinetic energy of the desorbing ion.

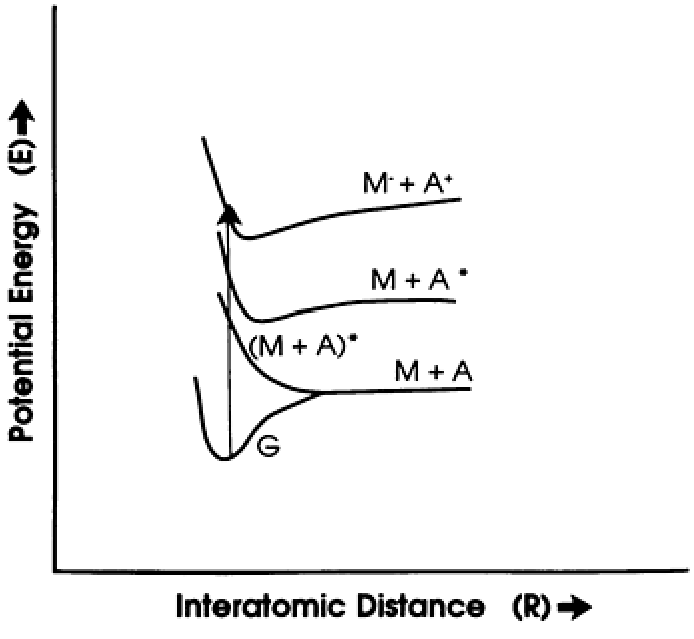


Fig. 24 Menzel-Gomer-Redhead (MGR) model of electron transition leading to ion desorption by ESD. Source: Ref 53

The KF model is characterized by an initial core-level ionization. This is followed by an interatomic electron transfer to fill the core-level vacancy, and this interatomic transition is also associated with the loss of at least one electron to the surroundings (vacuum). Note that the process described is an Auger process, and that interatomic, rather than intraatomic, electron transfer has been specified. This can occur when the atom in which the initial core-level vacancy occurs is present in an oxidation state such that no valence electrons are available to fill the core-level vacancy. An example is titanium in the +4 oxidation state. In this case, the metallic ion has been ionized to a noble gas configuration. This situation is shown in Fig. 25.

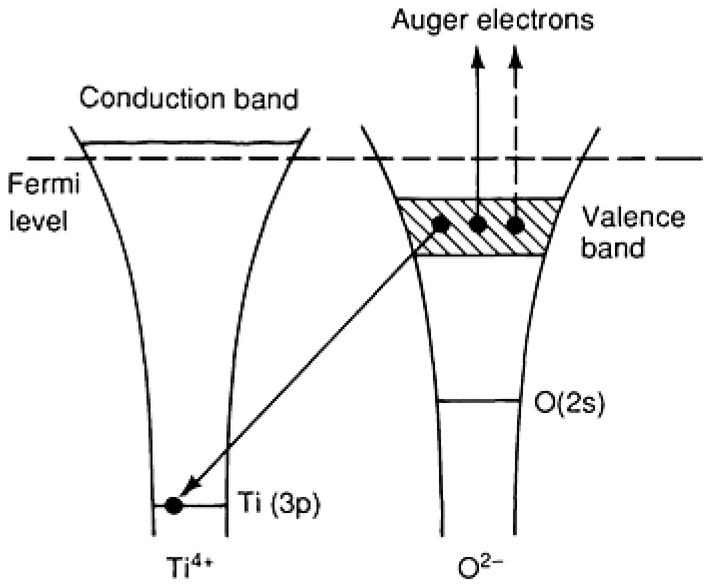


Fig. 25 Knotek-Feibelman (KF) model for ESD using TiO_2 as example. Depicted is interatomic electron transfer between oxygen and titanium. Source: Ref 54

ESD Equipment. The equipment required for ESD is analogous to that required for AES: an electron source, a method for detecting ion energies, and sufficient vacuum. ESD also requires a method of mass analysis to identify masses associated with specific energies. Historically, numerous types of system configurations have been employed for ESD studies. Quadruple mass analyzers are most commonly used.

However, ESD has been performed recently in conjunction with Auger and related spectroscopies. In this case, the electron gun utilized for AES is used for ESD, and the cylindrical mirror analyzer can be configured for ion energy analysis of positively charged ions (Ref 55). With this particular configuration, ion energies ranging from 0 to 20 eV can be detected. A disadvantage is that mass resolution is poor when the CMA is applied as a mass analyzer for species of mass greater than approximately 15 amu.

Traum and Woodruff (Ref 55) originally devised a method to operate the CMA as a time-of-flight mass analyzer and Gilbert *et al.* (Ref 56) have developed a pulse-counting circuit for operating the CMA in the time-of-flight regime. These articles are suggested as initial references for those interested in using the CMA analyzer as a time-of-flight mass analyzer.

ESD Spectrum. The measurement of the energy of ions being desorbed from a surface is sometimes referred to as an electron-stimulated desorption-ion energy distribution (ESDIED) spectrum, which represents the relative number of ions with a particular kinetic energy. An example spectrum for a cleaned polycrystalline tin oxide sample is shown in Fig. 26. Note that this spectrum appears to be the convolution of at least two peaks. At this point, time-of-flight analysis must be performed for each peak of interest in the ESDIED spectrum, so that chemical species associated with each peak can be identified. Peak values of 1.9 eV and 3.5 eV are assigned for the ESDIED spectrum shown in Fig. 26.

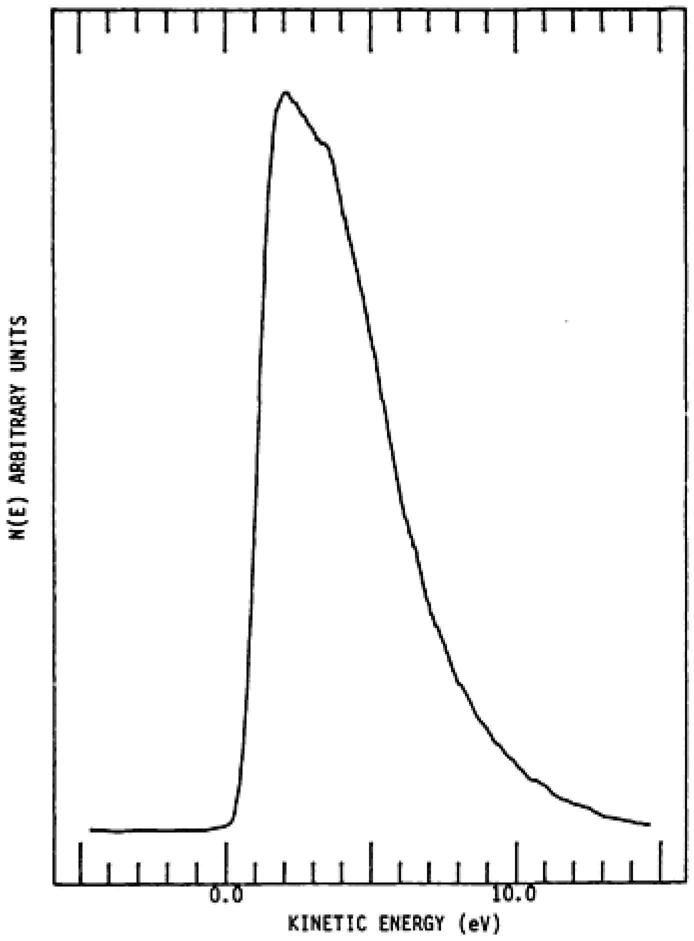


Fig. 26 ESDIED spectrum of cleaned polycrystalline tin oxide sample

Figures 27 and 28 contain time-of-flight data for the 1.9 and 3.5 eV ESDIED peaks, respectively. As shown in the spectrum, the lower-energy peak is associated with H^+, whereas the very small peaks at the higher energy end of the spectrum are associated with O^+ and OH^+. It should be noted that the presence of O^+ and/or OH^+ is not clear at this point because the signal is very low. However, it will be shown later that the assumption of some O^+ and/or OH^+ activity in this region is justified.

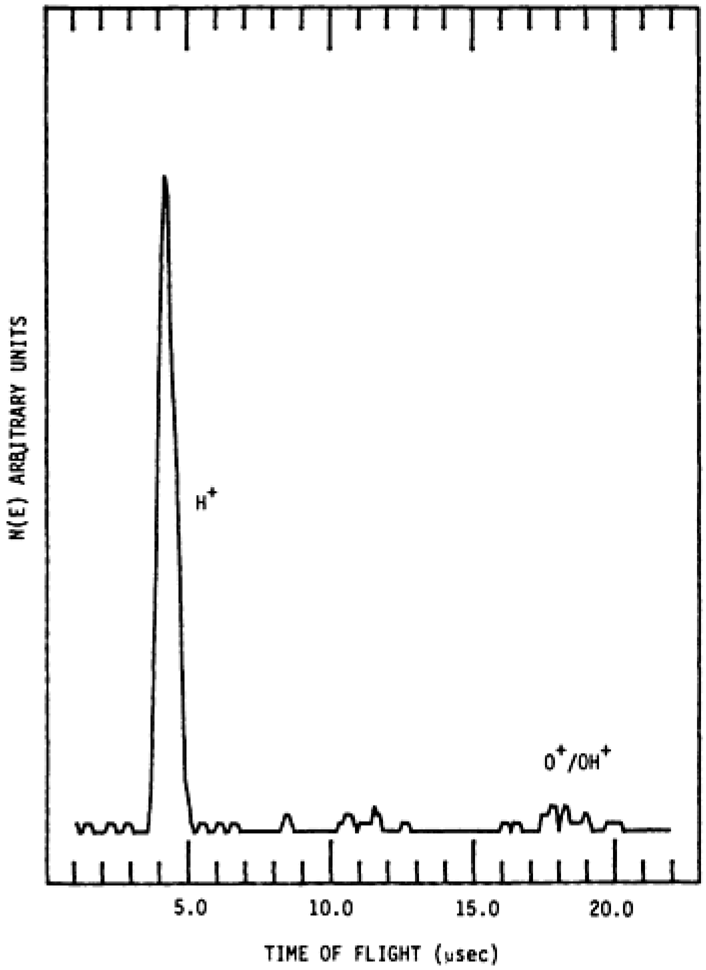


Fig. 27 Time-of-flight data for species composing the 1.9 eV ESDIED peak

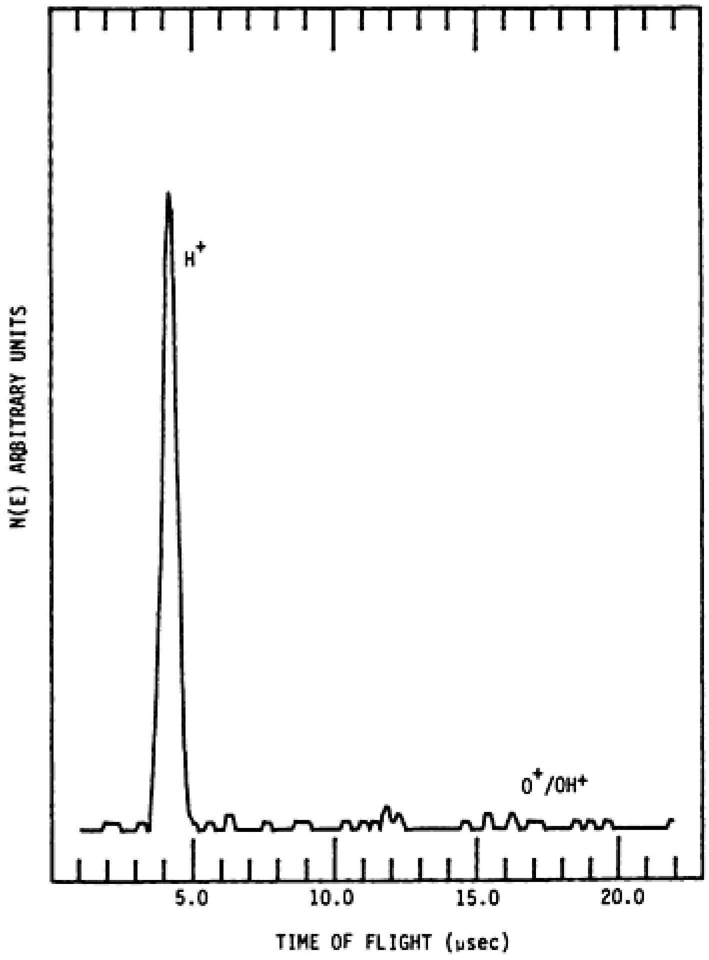


Fig. 28 Time-of-flight data for species composing the 3.5 eV ESDIED peak

It has been shown (Ref 55) that the time-of-flight of an ion is proportional to its mass. The relationship is time-of-flight = 4.2 $\sqrt{m}$, where *m* is the mass of the ion. Hence, flight times of 4.2, 16.4, and 16.9 μs are expected for H^+, O^+, and OH^+, respectively. Time-of-flight data for the 3.5 eV ESDIED peak also shows that H^+ is being desorbed, but that O^+ and/or OH^+ are probably not.

ESD Applications. The data acquired for the polycrystalline tin oxide sample discussed in the previous section is discussed further here. Consider the ESDIED and time-of-flight data shown in Fig. 29. The data were acquired after annealing for 30 min at 600 °C (1110 °F) and 9.3 × 10^{-7} Pa (7.0 × 10^{-9} torr). Note that the 1.9 eV ESDIED peak is very small relative to the higher-energy peak around 4.3 eV. The time-of-flight data of Fig. 30 and 31 show that O^+ and OH^+ are both desorbed, and the 4.3 eV peak has more O^+ than OH^+, relative to the 1.9 eV peak.

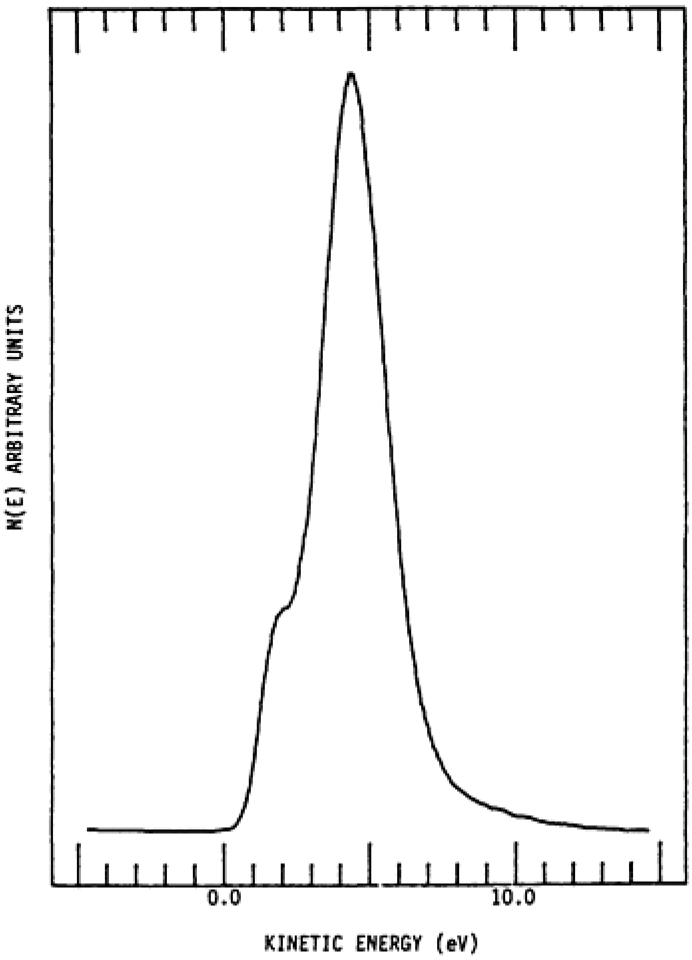


Fig. 29 ESDIED spectrum for polycrystalline tin oxide sample following annealing in vacuum for 30 min at 600 °C (1110 °F)

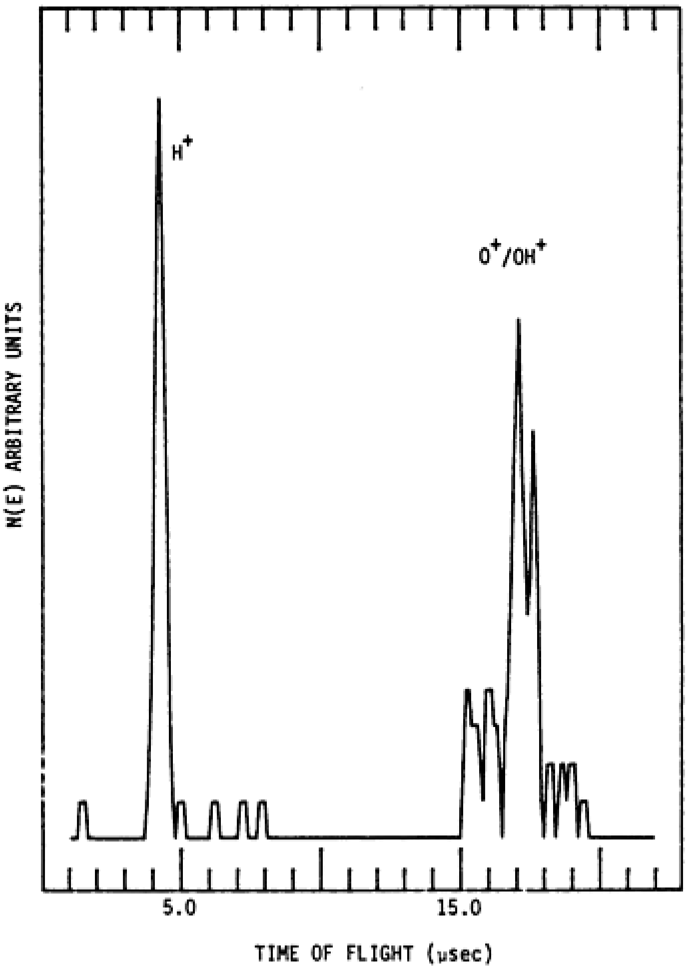


Fig. 30 Time-of-flight data for species composing the 1.9 eV ESDIED peak following annealing in vacuum for 30 min at 600 °C (1110 °F).

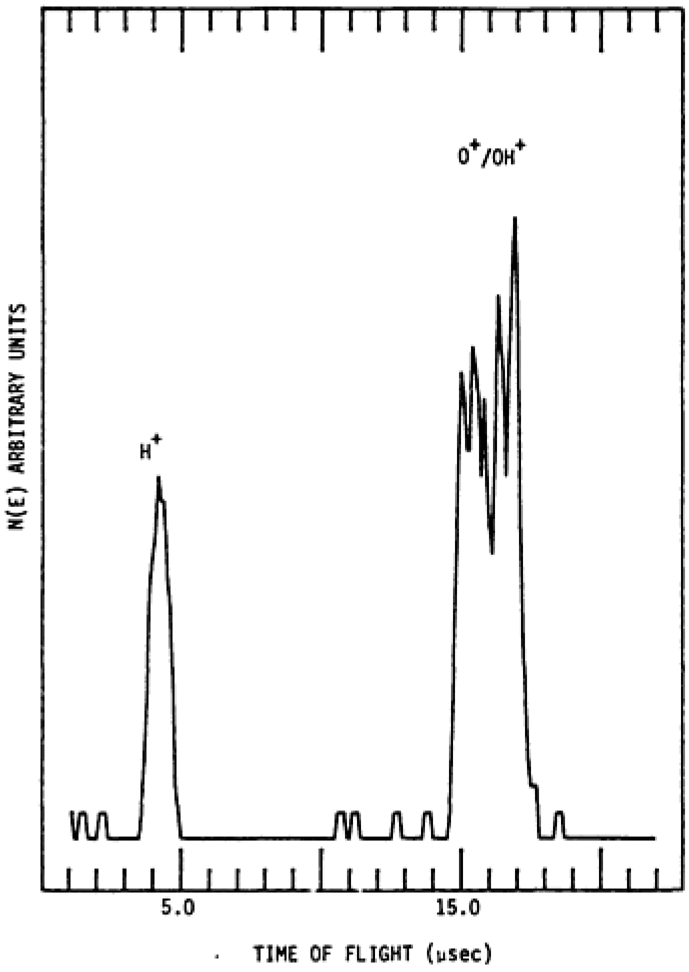


Fig. 31 Time-of-flight data for species composing the 4.3 eV ESDIED peak following annealing in vacuum for 30 min at 600 °C (1110 °F)

The ESD data can be explained by the following considerations. First, it is necessary to realize that tin oxide undergoes dehydration for the annealing conditions used. This has been shown by Cox (Ref 57) using valence band ESCA. If it is noted that the O^+ and OH^+ desorption signal is very small prior to annealing, but significantly larger following annealing, it is clear that the surface O^+ and OH^+ for the clean sample, prior to annealing, are not active with regard to desorption by electron stimulation. However, the O^+ and OH^+ that remain after dehydration are amenable to desorption by electron stimulation. Therefore, oxygen and hydrogen not associated with water of hydration have been distinguished because these bonding states are active with respect to ESD, whereas those associated with water of hydration are not.

It is possible to test this interpretation by exposing the sample to an oxidizing atmosphere. This was accomplished by annealing the sample in oxygen for 1.5 h at 400 °C (750 °F) at 1.3×10^{-4} Pa (10^{-6} torr.) The ESDIED and time-of-flight data are shown in Fig. 32 33 34. The 1.9 eV ESDIED peak has increased in size relative to that observed for the previous 600 °C (1110 °F) annealing treatment. The relative amounts of O^+ and OH^+ being desorbed have also decreased again, compared with data collected following the 600 °C (1110 °F) annealing treatment.

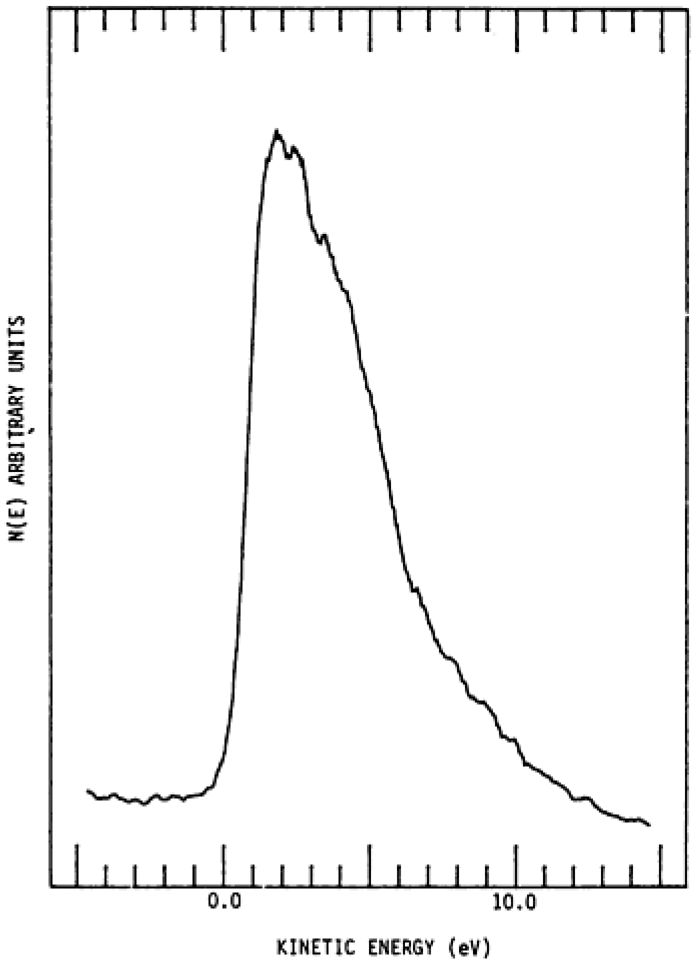


Fig. 32 ESDIED spectrum for polycrystalline tin oxide sample following annealing in oxygen for 90 min at 400 °C (750 °F)

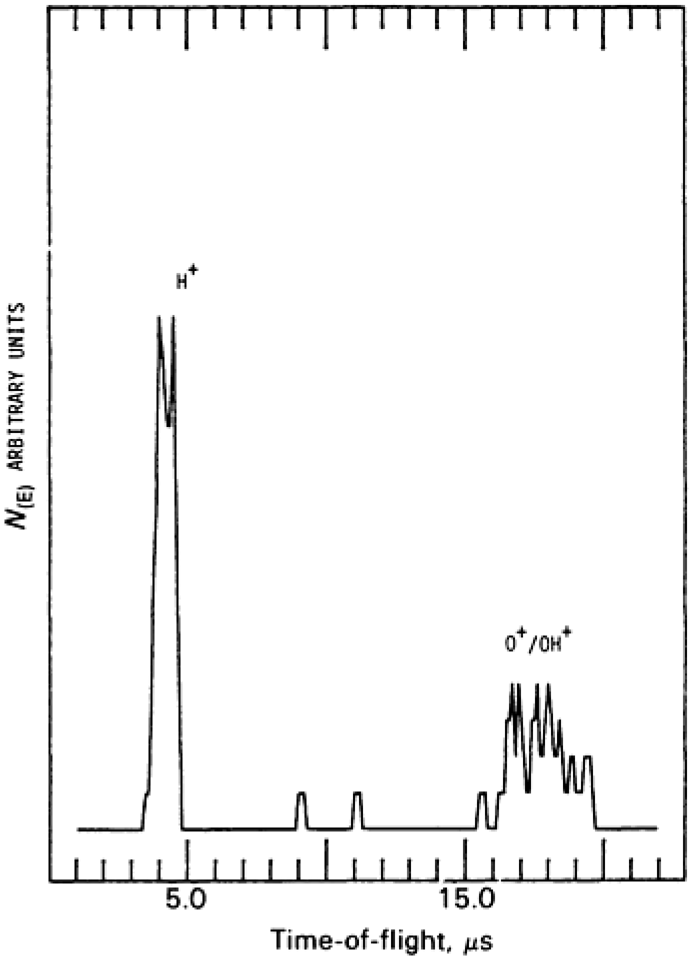


Fig. 33 Time-of-flight data for species composing the 4.3 eV ESDIED peak following annealing in oxygen for 90 min at 400 °C (750 °F)

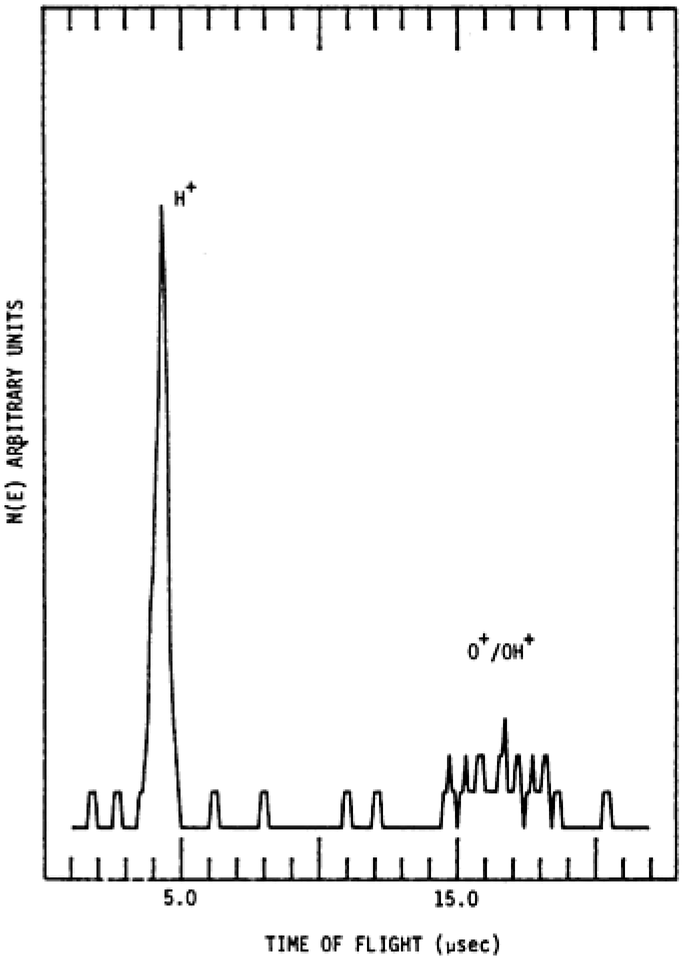


Fig. 34 Time-of-flight data for species composing the 4.3 eV ESDIED peak following annealing in oxygen for 90 min at 400 °C (750 °F)

It is reasonable to expect oxygen supplied to the sample during the oxidation treatment to become associated with vacancies generated during the dehydration process. Therefore, it can be concluded that the 1.9 eV ESDIED peak is associated with water of hydration, and the higher-energy ESDIED peak is associated with other species, such as oxygen bound to tin, in the lattice structure.

Thus, oxygen and hydroxyl groups associated with water of hydration are not active with regard to desorption by electron stimulation, relative to oxygen and hydroxyl groups associated with other bonding situations. ESD has been used to distinguish these oxygen and hydrogen types.

ESD is a valuable technique because it enables hydrogen detection and it can be used with minimum modifications to existing AES equipment. The major drawback is that data interpretation beyond identification of desorbing species is often difficult. An alternative to ESD is discussed next.

Secondary Ion Mass Spectrometry (SIMS)

SIMS is an analytical technique that has become very popular over the past few years. Enhanced element sensitivity and hydrogen detection capability (order of ppm) are two advantages of SIMS that AES and XPS techniques do not offer. Its primary disadvantages are that it is inherently a destructive technique and quantitation is more difficult, relative to techniques such as AES and XPS.

Specific examples of the application of SIMS to studies of wear, lubrication, and friction are somewhat limited, when compared with techniques such as AES. An excellent example of the application of SIMS to study material transfer

resulting from abrasive contact between a ceramic and several metals is described in Ref 58. SIMS is particularly suited to such studies if the amounts of material transferred are expected to be very small.

SIMS Fundamentals. The SIMS process is performed by bombarding the surface of a solid target material of interest with a beam of energetic ions. The ions composing the bombarding beam are referred to as primary ions. These ions can be delivered to the target surface at energies up to approximately 40 keV. The result of collision processes between primary ions and the target surface of interest is the emission of negative, positive, and neutral species. The term "species" is employed to indicate that ions or agglomerations of atoms bearing a net charge can be emitted from the surface. The species emitted from the surface are analyzed in terms of their mass-to-charge ratios (m/e). Therefore, only charged species can be analyzed. Neutral species that have been sputtered from the surface must first be ionized before analysis is possible.

It is important to point out that charged species leaving the target surface as a result of sputtering constitute only a small portion of all sputtered atoms leaving the surface. This typically ranges from a few hundredths of a percent, up to approximately 1%. Discussions of analytical descriptions of secondary ion yield and parameters that influence the overall yield, such as ionization probability, are discussed in Ref 46 and 47.

SIMS can be performed in two modes. In one case, the primary ion beam is rastered over the surface covering an area of approximately 50 × 50 μm (2 × 2 mils) (Ref 46). This mode of analysis is referred to as static SIMS and is characterized by a relatively slow removal rate of atoms from the surface of the target material.

Alternatively, the primary ion beam can be focused on an area of submicron dimension and material removed at very high rates relative to static SIMS. This mode of operation is referred to as dynamic SIMS. Its sampling depth is on the order of 10^3 nm, whereas static SIMS is characterized by sampling depths of only a couple of atomic layers (Ref 46).

SIMS is often employed to obtain chemical depth profile information. Dynamic SIMS can achieve practical sputter rates, as well as maximum sensitivity (Ref 46). Therefore, this section focuses on this mode.

In most cases, the primary ion beam employed with SIMS is an inert gas. However, primary beam systems should be capable of generating both negatively and positively charged ions of reactive gases. Negatively charged ions, such as O^-, can be used as a primary beam source when sample charging is expected to be a problem. With versatility in the primary beam source, electrically insulating materials can be analyzed with a minimum of difficulties.

An additional advantage of the ability to implement several types of primary ion beam gases is the effect of secondary ion yield. This is because secondary ion yields are influenced by the charge characteristics at the surface. Ion yields are maximized when neutralization probabilities are low. Therefore, positively charged secondary species are less likely to be neutralized when electronegative atoms are present in surface and near-surface regions of the material from which secondary ions are being sputtered. For this reason, if oxygen is used as a primary beam source, rather than an inert gas, the emission of positively charged secondary ions can be expected to be enhanced. This point is discussed in Ref 47.

The SIMS spectrum consists of a representation of the signal intensity, or count rate, as a function of mass-to-charge ratio (m/e). Consider the SIMS data presented in Fig. 35. These data represent a SIMS survey scan for the surface and near-surface regions of a BiO_x-Au-glass thin-film system, such as that discussed in the AES section (see Fig. 17). The SIMS data in Fig. 35 contain several elements that do not appear in the film data in Fig. 17 (that is, elements in addition to Bi, O, Au, and Si). These additional elements are the result of trace contamination of the deposition chamber from previous deposition processes. This is an excellent example of the increased sensitivity of SIMS relative to AES.

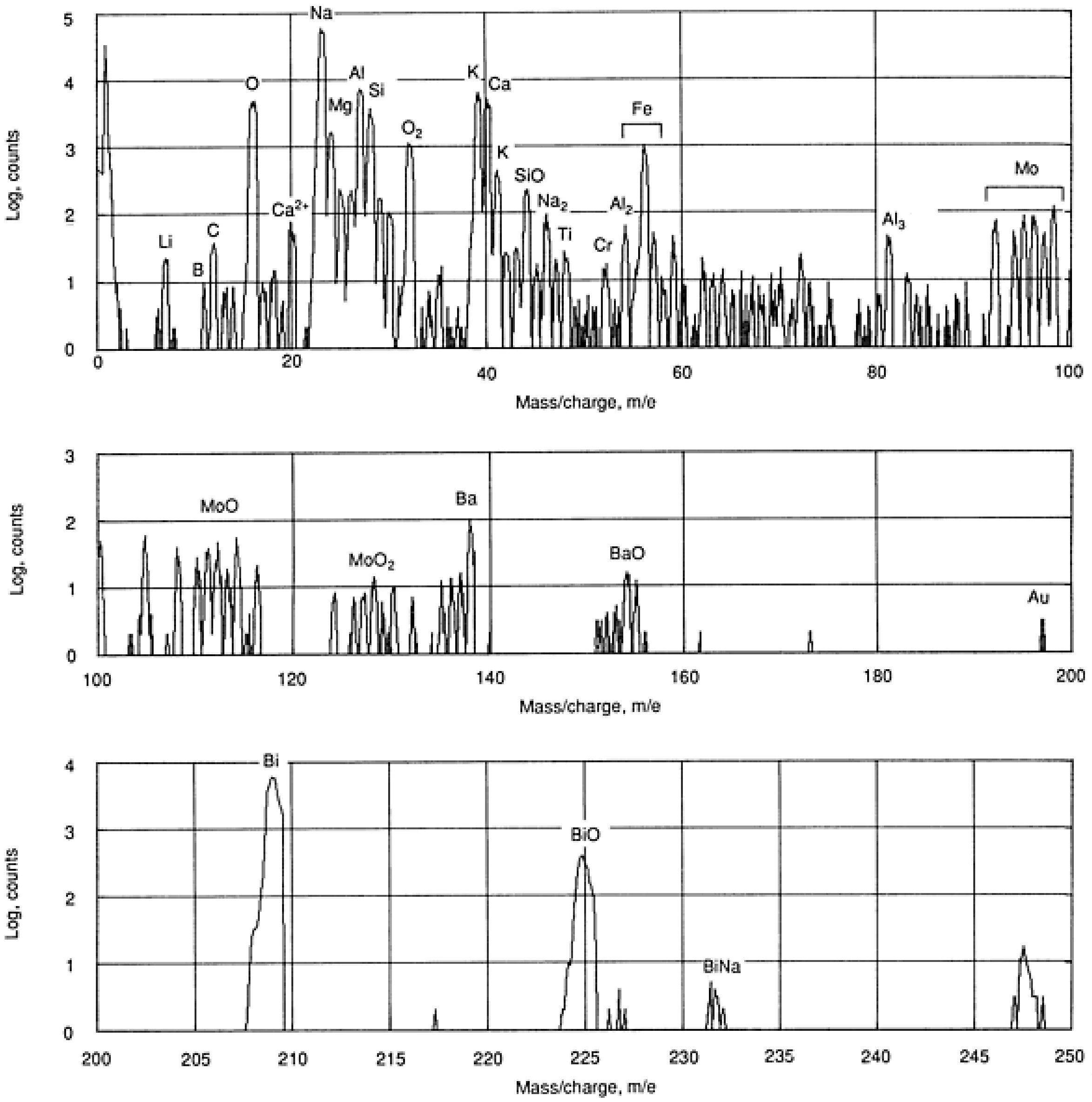


Fig. 35 SIMS survey of bismuth layer of BiO_x-Au-glass specimen

The reader should be aware that quantitation of SIMS data is a very complicated issue. The basis problem is the fact that the secondary ion yield can be significantly influenced by several phenomena that can be very complex in nature. The emission of secondary ions depends on the chemical and physical characteristics of the target surface, the primary beam characteristics, and the matrix characteristics of the target material.

The most direct method of quantitation is to compare SIMS results with a reference material of known composition for a given set of conditions for data acquisition. Of course, this involves the availability of reference materials for all anticipated matrix configurations that might be encountered. At best, this is inconvenient. It would therefore be helpful if a myriad of matrix configurations could be evaluated by simply extending data required for a minimum number of reference samples.

In principle, this could be accomplished by the ability to extend relative sensitivity factors obtained for a few reference materials to any arbitrary matrix configuration. This has been achieved by defining a parameter that characterizes the electronic properties of the surface of the target material (Ref 59).

It is of interest to note that data from internal standards have been employed to specify parameters of a model to predict relative atomic fraction values from secondary ion intensities (Ref 60). The model referenced here has been applied with a reasonable degree of success when one considers that it has been applied to a variety of materials.

SIMS equipment includes an ion source (ion gun), a UHV environment, and a detection system that consists of components such that energy and mass selection of the sputtered species occurs prior to the detector. References 47 and 59 provide a more detailed equipment description.

SIMS Applications. As previously mentioned, SIMS is a valuable technique in terms of detecting elements present in relatively small concentrations, as well as detecting hydrogen. These capabilities can be particularly useful in friction, lubrication, and wear problems involving materials that contain hydrogen and/or where small amounts of material transfer occur.

Material transfer examples include that which results when materials are in tribocontact and that which results from diffusion processes. Material transfer by diffusion is often difficult to substantiate, especially if the amount of mass transfer is relatively small.

For the purpose of demonstration, the interdiffusion of constituents composing a layered BiO_x-Al-glass system is discussed. This system is analogous to the BiO_x-Au-glass thin-film system previously discussed and is used here because the amount of interdiffusion associated with high-temperature oxidation of the BiAl-glass-layered structure is substantial and can be clearly identified.

Consider the SIMS depth profile data shown in Fig. 36 and 37. These data respectively represent results for a layered structure that was not subjected to high-temperature oxidation, and a layered structure that was. The secondary ions monitored are Al^+, Bi^+, O^+, and Si^+. The data clearly demonstrate that interdiffusion occurs among constituents of the bismuth layer, the metal layer, and the glass substrate. It is interesting to note that interdiffusion processes for this system were, in general, not nearly as observable with AES depth profiling as with SIMS. Again, this demonstrates the advantage of the increased sensitivity of SIMS, relative to other surface analytical techniques.

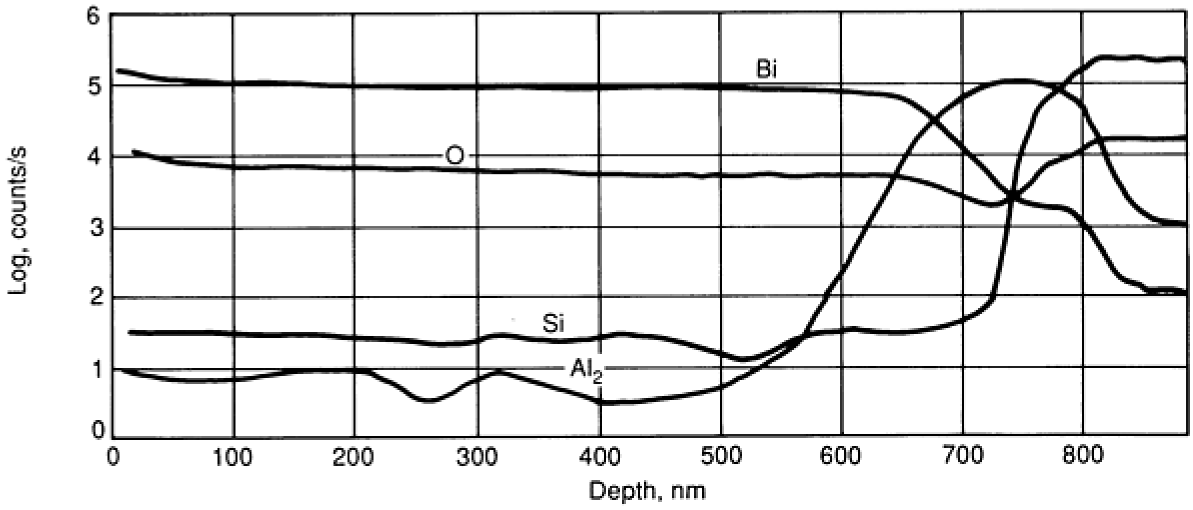


Fig. 36 SIMS depth profile of layered Bi-Au-glass specimen

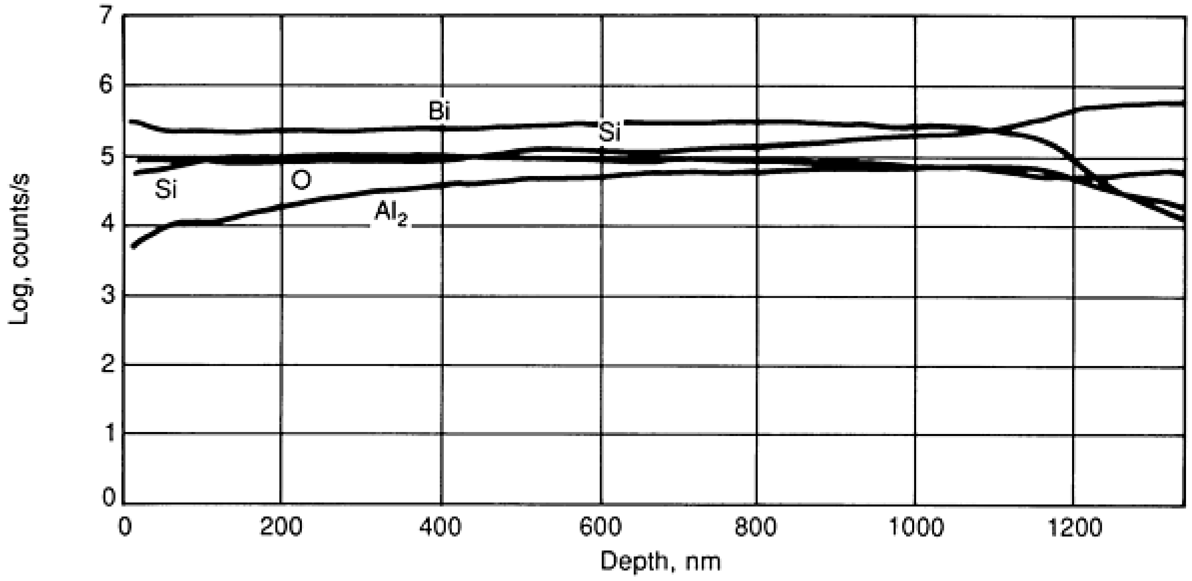


Fig. 37 SIMS depth profile of layered BiO_x-Au-glass specimen

It is important at this point to emphasize that care must be taken when evaluating data acquired near interfacial regions. This consideration is often applicable when mass transport by diffusion is of interest. The important point is that the sputtering process can produce mixing of constituents across the interface, which results in peak broadening in the composition profile. Such broadening can be interpreted as diffusion effects if sufficient care is not taken. The most practical way to avoid difficulty is to evaluate reference materials for comparison, as was done here by comparing a sample not subjected to high-temperature oxidation to a sample that was.

The capability to detect hydrogen can be very useful in friction, lubrication, and wear studies because such studies often involve systems that contain hydrogen. For instance, Sugita and Ueda (Ref 61) studied the wear characteristics of silicon nitride in water with respect to the material produced at the specimen-counterface interface. SIMS analysis showed that the worn surface contained silicon bonded to oxygen and hydrogen. Therefore, silicon was oxidized as a result of the tribocontact. The presence of hydrogen suggests that the oxidized silicon is hydrated to some extent. X-ray diffraction data were used in conjunction with the SIMS data to propose a mechanism of material removal for this system of silicon nitride rubbing against silicon nitride in a water environment. The mechanism proposed is one in which the silicon in silicon nitride is first oxidized and then converted to an amorphous form of silica hydrate, which is then removed by frictional forces associated with the rubbing of the two silicon nitride faces.

Thus, SIMS provides a means by which hydrogen can be detected and a means by which elements can be observed at very low concentrations. Its primary disadvantages are that it is inherently a destructive technique and chemical bonding information such as that obtained with XPS and AES is not available.

Infrared (IR) Spectroscopy

The IR spectroscopy technique should be mentioned because it is becoming more important in the study of the chemistry of solid surfaces; however, it will not be treated in detail in this article. Infrared spectroscopy has been employed for some time as a routine technique for determining the molecular structure of organic compounds, and therefore it is not a new technique.

IR spectroscopy is performed by subjecting the sample to a source of IR radiation. This source is sometimes referred to as an emitter. For practical purposes, the IR range is taken to be electromagnetic radiation within the energy range for 200 to 4000/cm. In practice, more than one type of emitter is required to cover the entire IR range. The electric field of the electromagnetic radiation can couple with oscillating dipoles of vibrating molecules. The result of this interaction, or coupling, of the electromagnetic radiation with vibrational energy modes of molecules is absorption of the radiation.

The IR absorption spectrum appears in the form of the percent of radiation transmitted through the sample (that is, not absorbed) as a function of IR radiation energy. Structures that are expected to be active with regard to IR absorption are those that exhibit a net dipole moment. Symmetric molecules, such as N_2 and O_2, are not expected to exhibit an absorption species for any relative position of the atoms (bond length). On the other hand, molecules such as HCl, NO, and CO are expected to exhibit characteristic absorption frequencies.

Problems in friction, lubrication, and wear can involve analysis of surface components that are present in coverages of considerably less than one monolayer. This is possible with IR spectroscopy, but most investigators employ the technique of multiple internal reflectance (MIR) spectroscopy for such studies. This technique offers increased sensitivity for surface components, allowing detailed studies of adsorbate-surface interactions. In fact, MIR spectroscopy can be used to determine the orientation of adsorbate molecules on solid surfaces. Both IR and MIR spectroscopy are discussed further in Ref 62 and 63.

References

1. H. Hertz, *Ann. Physik,* Vol 31, 1887, p 983
2. J.J. Thompson, *Phil. Mag.,* Vol 48, 1899, p 547
3. A. Einstein, *Ann. Physik,* Vol 17, 1905, p 132
4. K. Siegbahn, C. Nordling, G. Johansson, J. Hedman, R.-F. Heden, K. Hamrin, U. Gelius, T. Bergmark, L.O. Werme, R. Manne, and Y. Baier, *ESCA Applied to Free Molecules,* North-Holland, Amsterdam, 1969
5. K. Siegbahn, C. Nordling, A. Fahlman, R. Nordberg, K. Hamrin, J. Hedman, G. Johansson, R. Bergmark, S.-E. Karlsson, I. Lindgren, and B. Lindberg, ESCA: Atomic, Molecular and Solid State Structures Studied by Means of Electron Spectroscopy, *Nova Acta Regiae Soc. Sci. Upsaliensis,* Ser. IV, Vol 20, Almquist and Wiksells, Uppsala, 1967
6. T.A. Carlson, *X-Ray photoelectron Spectroscopy,* T.A. Carlson, Ed., Dowden, Huntington & Ross, Inc., 1978
7. K.W. Nebesny, B.L. Maschhoff, and N.R. Armstrong, *Anal. Chem.,* Vol 61, 1989, p 469A
8. A.A. Galuska, *J. Vac. Sci. Technol. B,* Vol 8, 1990, p 488
9. S.D. Gardner, G.B. Hoflund, M.R. Davidson, and D.R. Schryer, *J. Catalysis,* Vol 115, 1989, p 132
10. M.R. Davidson, G.B. Hoflund, L. Niinista, and H.A. Laitinen, *J. Electroanal. Chem.,* Vol 228, 1987, p 471
11. G.B. Hoflund, D.A. Asbury, S.J. Babb, A.L. Grogan, Jr., H.A. Laitinen, and S. Hoshino, *J. Vac. Sci, Technol. A,* Vol 4, 1986, p 26
12. G.B. Hoflund, A.L. Grogan, Jr., D.A. Asbury, H.A. Laitinen, and S. Hoshino, *Appl. Surf. Sci.,* Vol 28, 1987, p 224
13. G.B. Hoflund, M. Davidson, E. Yngvadottir, H.A. Laitinen, and S. Hoshino, *Chem. Mater.,* Vol 1, 1989, p 625
14. M.R. Davidson, G.B. Hoflund, and R.A. Outlaw, *J. Vac. Sci. Technol. A,* Vol 9, 1991, p 1344
15. M. Cardona and L. Ley, *Photoemission in Solids I: General Principles,* M. Cardona and L. Ley, Ed., Springer-Verlag, Berlin, 1978
16. A. Azouz and D.M. Rowson, A Comparison of Techniques for Surface Analysis of Extreme Pressure Films Formed During Wear Tests, *Microscopic Aspects of Adhesion and Lubrication,* J.M. Georges, Ed., Elsevier, Amsterdam, 1982
17. B.A. Baldwin, *Lubr. Eng.,* Vol 32, 1976, p 125
18. R.J. Bird, *Wear,* Vol 37, 1976, p 132
19. G.B. Hoflund, H.-L. Yin, A.L. Grogan, Jr., D.A. Asbury, H. Yoneyama, O. Ikeda, and H. Tamura, *Langmuir,* Vol 4, 1988, p 346
20. C.R. Brundle and A.D. Baker, Ed., *Electron Spectroscopy: Theory, Techniques and Applications,* Academic Press, 1977
21. C.D. Wagner, W.M. Riggs, L.E. Davis, J.F. Moulder, and G.E. Muilenberg, Ed., *Handbook of X-Ray Photoelectron Spectroscopy,* Perkin-Elmer Corporation, 1979

22. S. Hagstrom, C. Nordling, and K. Siegbahn, *Z. Phys.,* Vol 178, 1964, p 439
23. J.A. Gardella, Jr., *Anal. Chem.,* Vol 61, 1989, p 589A
24. D.M. Hercules, *Anal. Chem.,* Vol 50, 1978, p 743A
25. H. Fellner-Feldegg, U. Gelius, B. Wannberg, A.G. Nilsson, E. Basilier, and K. Siegbahn, *J. Electron Spectrosc. Relat. Phenom.,* Vol 5, 1974, p 643
26. P.W. Palmberg, *J. Electron Spectrosc.,* Vol 5, 1974, p 691
27. G.B. Hoflund, D.A. Asbury, C.F. Corallo, and G.R. Corallo, *J. Vac. Sci. Technol.,* Vol 6, 1988, p 70
28. H. Ferber, C.K. Mount, G.B. Hoflund, and S. Hoshino, Surface Studies of N Implanted and Annealed ABCD Chromium Films, *Thin Solid Films,* Vol 203, 1991, p 121
29. V.D. Castro and G. Polzonetti, *J. Electron Spectrosc. Relat. Phenom.,* Vol 48, 1989, p 117
30. D.P. Smith, *Surf. Sci.,* Vol 25, 1971, p 171
31. E.P.Th.M. Suurmeijer and A.L. Boers, *Surf. Sci.,* Vol 43, 1973, p 309
32. E. Taglauer and W. Heiland, *Appl. Phys.,* Vol 9, 1976, p 261
33. E. Taglauer and W. Heiland, *Surf. Sci.,* Vol 33, 1972, p 27
34. S.H.A. Bageman and A.L. Boers, *Surf. Sci.,* Vol 30, 1972, p 134
35. D.S. Karpuzov and V.E. Yurasova, *Phys. Status Solidi (b),* Vol 47, 1971, p 41
36. R.F. Goff and D.P. Smith, *J. Vac. Sci. Technol.,* Vol 7, 1970, p 72
37. W. Heiland, H.G. Schäffler, and E. Taglauer, *Surf. Sci.,* Vol 35, 1973, p 381
38. X.Z. Jiang, T.F. Hayden, and J.A. Dumesic, *J. Catalysis,* Vol 83, 1983, p 168
39. A.J. Simoens, R.T. Baker, D.J. Dwyer, C.R Lund, and R.J. Madon, *J. Catalysis,* Vol 86, 1984, p 359
40. P.N. Belton, Y.M. Sun, and J.M. White, *J. Phys. Chem.,* Vol 8, 1984, p 5172
41. S.J. Tauster, S.C. Fung, and R.L. Gartner, *J. Am. Chem. Soc.,* Vol 106, 1980, p 170
42. H. Niehus and E. Bauer, *Surf. Sci.,* Vol 47, 1975, p 222
43. S.V. Pepper, *J. Appl. Phys.,* Vol 45, 1974, p 2947
44. N. Takahashi and K. Okador, *Wear,* Vol 38, 1976, p 177
45. H.J. Mathien and D. Landolt, *Wear,* Vol 66, 1981, p 87
46. D. Briggs and M.P. Seah, *Practical Surface Analysis by Auger and X-Ray Photoelectron Spectroscopy,* John Wiley & Sons, 1983
47. L.C. Feldman and J.W. Mayer, *Fundamentals of Surface and Thin Film Analysis,* North-Holland, 1986
48. C. Kittel, *Introduction to Solid State Physics,* 5th ed., John Wiley & Sons, 1976
49. L.E. Davis, N.C. MacDonald, P.W. Palmberg, G.E. Riach, and R.E. Weber, *Handbook of Auger Electron Spectroscopy,* Physical Electronics Industries, Inc., 1976
50. G.B. Hoflund, A.L. Grogan, Jr., and D.A. Asbury, *J. Catalysis,* Vol 109, 1988, p 226
51. A.L. Grogan, Jr., V.H. Desai, S.L. Rice, and F. Gray III, Apparatus for Chemomechanical Wear Studies with Biaxial Load and Surface Charge Control, *Wear,* accepted for publication
52. G.B. Hoflund, *Scanning Electron Microsc.,* Vol IV, 1985, p 1391
53. D. Menzel and Gomer, *J. Chem. Phys.,* Vol 41, 1964, p 3311
54. M.L. Knotek and P.J. Feibelman, *Phys. Rev. Lett.,* Vol 40, 1978, p 964
55. M.M. Traum and D.P. Woodruff, *J. Vac. Sci. Technol. A,* Vol 17, 1980, p 1203
56. R.E. Gilbert, D.F. Cox, and G.B. Hoflund, *Rev. Sci. Instrum.,* Vol 53, 1982, p 1281
57. D.F. Cox, G.B. Hoflund, and H.A. Laitinen, *Appl. Surf. Sci.,* Vol 20, 1984, p 30
58. K. Fujiwara, *Wear,* Vol 51, 1978, p 127
59. A.W. Czanderna, Methods of Surface Analysis, *Methods and Phenomena, Their Applications in Science and Technology,* Elsevier, Amsterdam, 1975
60. C.A. Anderson and J.R. Hinthrone, *Anal. Chem.,* Vol 45, 1973, p 1421
61. T. Sugita and K. Ueda, *Wear,* Vol 97, 1984

62. B.P. Straughan and S. Walton, Ed., *Spectroscopy,* Chapman and Hall, London, 1976
63. N.J. Harrick, *Internal Reflection Spectroscopy,* Interscience, 1967

X-Ray Characterization of Surface Wear

C.R. Houska, Virginia Polytechnic Institute and State University

Introduction

X-RAY DIFFRACTION and spectroscopy provide a variable-depth probe of the atomic arrangements and composition of near-surface material. Depending on the sample, x-ray wavelength, and experimental arrangement, quantitative data from about 10 nm to a number of μm can be provided. This range of penetration depth can be used to examine wear-modified and unmodified regions. Deeper penetrations allow unmodified material to be sampled as a reference for those changes taking place near the surface.

The first consideration is to examine the conditions that determine penetration depths for both x-ray diffraction (XRD) and x-ray fluorescence (XRF). Both kinds of data can be obtained using either commercially available equipment or specialized beamlines at synchrotron radiation facilities. Synchrotron beamlines allow more flexibility in the choice of wavelengths and provide highly collimated and intense beams that add to the capability to alter the penetration distance by adjusting the optical arrangement. Small glancing angles between the specimen surface and either the incident or diffracted beams are effective in probing closer to the surface.

X-rays can become totally reflected from locally flat surfaces at extremely low angles of incidence (<0.25°). Under these conditions, beam penetration becomes anomalous and limited to less than 8 nm. Penetration distance, Z_0^T, depends on the angle of incidence, θ_i, relative to the critical angle, θ_c; that is,

$$Z_0^T \approx \frac{\lambda}{2\pi(\theta_c^2 - \theta_i^2)} \qquad \textbf{(Eq 1)}$$

where π = x-ray wavelength (Ref 1).

An uncertainty or divergence in the angle of the incident beam, θ_i, can introduce a large uncertainty in penetration distance. A wavy or rough surface and an ideally parallel beam also leads to quantitative uncertainties because θ_i takes on a range of values. Because of these difficulties in determining penetration distance with total reflection, the subsequent discussion involves glancing angles that are greater than the critical angle. Above the critical angle, one can define penetration distances more simply using a large data file of absorption coefficients (Ref 2) and more conventional equipment.

Surface roughness introduces an additional complication. Near the surface, signal-producing material is removed. Where it exist, the beam paths differ from point to point at a fixed distance below the mean surface. A fluctuation in path length produces a net decrease in the measured signal (Ref 3). This problem has been studied using severely ground samples and a correction given in terms of a Gaussian distribution of asperities with correlation (Ref 4).

X-Ray Diffraction and Fluorescence from Flat Surfaces

Analyses of penetration distances are typically based on flat surfaces. The surface need only be flat over a distance that allows the incident and outgoing signal beams to see a locally flat region of the surface. Under these conditions, beam paths enter and leave near-surface material only once. Both diffraction and fluorescence signals are treated together with a common absorption term. XRF has one more absorption term, because of a difference in the incident and fluorescence wavelengths.

The following discussion refers to a differential element of irradiated area

$$A_e = \frac{A_0}{\sin(\theta + \omega)} \quad \textbf{(Eq 2)}$$

and thickness, d*Z*. The various terms, along with the tilt angle, ω, are illustrated in Fig. 1. A_0 is the cross section of the incident beam with incident angle, $\theta + \omega$, and intensity, I_0. If the signal angle is $\theta - \omega$, the fluorescent signal from an element at a depth, *Z*, is given by (Ref 5):

$$dP = I_0 Q_F \exp\left[\frac{-\mu_i Z}{\sin(\theta + \omega)} - \frac{\mu_s Z}{\sin(\theta - \omega)}\right] dV \quad \textbf{(Eq 3)}$$

with

$$dV = \frac{A_0\, dZ}{\sin(\theta + \omega)} \quad \textbf{(Eq 4)}$$

and

$$Q_F = \frac{N_i\, \sigma_i\, D_i}{4\pi R^2} \quad \textbf{(Eq 5)}$$

where N_i = number of *i* atoms/volume, σ_i = total atomic cross section of atom *i*, D_i = dimensionless absorption factor for fluorescent radiation in the path from sample surface to detector multiplied by the detector efficiency, and *R* = sample-to-detector distance.

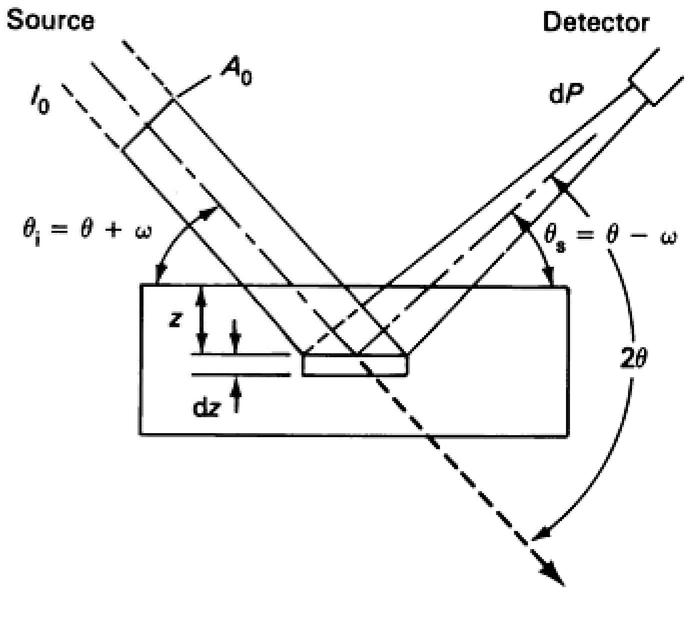


Fig. 1 X-ray optics illustrating: θ_i = angle of incidence, θ_s = signal angle, ω= angle of tilt from symmetrical arrangement, and 2θ = angle between signal and incident beam. Other quantities include A_0 = cross-sectional area of incident beam, and total volume of signal element is A_0/sin(θ + ω) multiplied by its thickness, *dZ*, located at a depth - *Z*. The detector is located at a distance *R* from the sample.

For a diffracting element, Q_D differs, depending on whether the volume dV contains a single crystal or a polycrystalline substance. Only a polycrystalline substance with a flat surface, which gives (Ref 6, 7)

$$Q_D = const.\ \frac{\lambda^3}{V_C^2} \left(\frac{1 + \cos^2 2\theta' \cos^2 2\theta}{(1 + \cos^2 2\theta') \sin\theta \sin 2\theta} \right) j\, F^2 \qquad \textbf{(Eq 6)}$$

will be considered. Here, θ is the Bragg angle for the sample, V_C is the volume of the unit cell, θ' is the Bragg angle of the monochromator, if used, j is the multiplicity, and F is the structure factor. If the thermal factor is known, it can be included as an additional term in Q_D.

The exponential term in Eq 2 is of primary interest because it determines beam penetration. With a fluorescence signal, two linear absorption coefficients are required; that is, one for the incident beam, μ_i, and another for the signal as it leaves the sample, μ_s. For diffraction, the incident and signal wavelength are the same with $\mu_i = \mu_s = \mu$. The absorption coefficient depends on the material and wavelength.

With a homogeneous sample, the contributions from all layers are obtained by integrating Eq 3. One normally integrates to infinity for a thick sample, giving the effective penetration distance.

$$Z_0 = (\mu_i S)^{-1} \qquad \textbf{(Eq 7)}$$

where the path length factor is

$$S = \frac{1}{\sin(\theta + \omega)} + \left(\frac{\mu_s}{\mu_i} \right) \frac{1}{\sin(\theta - \omega)} \qquad \textbf{(Eq 8)}$$

The total signal for all elements becomes

$$\frac{P}{I_0 Q_i} = \frac{A_0 (\mu_i S)^{-1}}{\sin(\theta + \omega)},\ i = F \text{ or } D \qquad \textbf{(Eq 9)}$$

which is written so as to isolate the effective volume term on the right. The term on the left is the reduced integrated intensity, where P is determined by the total area under a peak.

If one considers a pair of rays entering and leaving the sample, penetrating to a distance equal to the effective penetration distance, Z_0, the beam is reduced by e^{-1} purely from sample absorption. The accumulated signal from all elements between the surface and Z_0 relative to an infinitely thick sample is $1 - e^{-1}$, or 0.63. As λ increases, the trend is for the penetration distance to decrease as the absorption coefficient usually increases. However, crossing an absorption edge causes a sudden change in the penetration distance (Ref 8). Another common way of decreasing the penetration distance is to decrease the angle of incidence below the Bragg angle, θ, with a fixed 2θ. This decreases S^{-1} for a particular Bragg peak as one goes toward smaller glancing angles. In fact, one can tilt the sample so that the glancing angle is one the signal side and obtain a similar decrease. The direction of tilt does influence the area of beam as it intersects with the sample and gives a higher effective volume at low angles of incidence and, therefore, more intensity.

Figure 2 illustrates changes in the effective penetration for a large in ω angles with CuKα (λ = 0.1542 nm), CrKα (λ = 0.2291 nm), a partially stabilized zirconia sample (PSZ), and the (111) and (400) Bragg reflections (Ref 9). A wavelength of 0.4 nm is also shown with the (111) set to illustrate an extreme wavelength that is only available at a synchrotron radiation beamline. Here, it can be seen that when the angle of incidence equals the signal angle, the effective penetration is a maximum at 0.42 μm, and can be further reduced to 0.02 μm for ± ω tilts of ~40°. This would give glancing angles below 2.9° for a Bragg angle of 42.9°. In other words, 63% of the Bragg peak would come from the region between the

surface and 0.02 μm. Changing λ from CrKα to CuKα, radiation increase the maximum penetration from 1.05 to 2.08 μm.

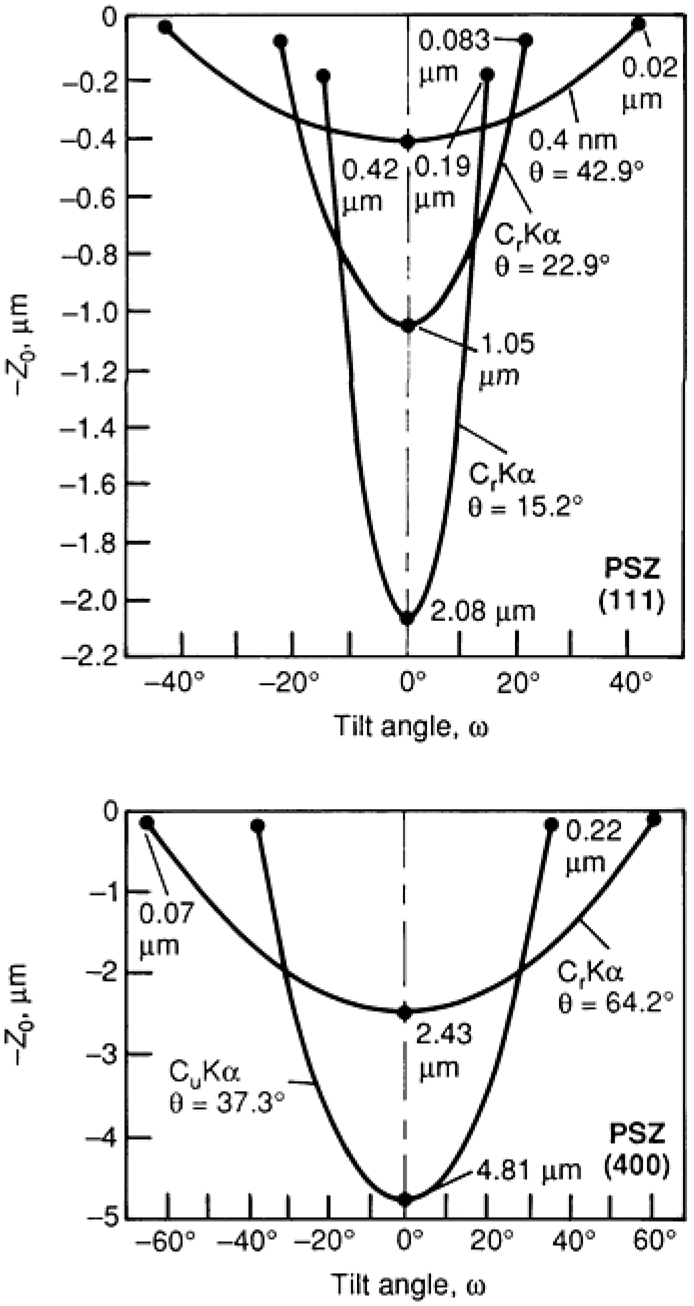


Fig. 2 Examples of effective flat sample penetration depths for (111) and (400) peaks with various wavelengths (0.4 nm, CrKα) and CuKα and tilt angle ω. All are based on the absorption in partially stabilized zirconia and its lattice constant. The maximum ω tilt is limited by θ and the critical angle for total reflection.

One would have to go to even lower glancing angles to attain 0.02 μm, which may not be practical with commercial x-ray systems. The relatively high absorption coefficient of zirconia tends to give limited penetration. A similar family of curves is found for the (400), but with larger penetration distance, because of the larger Bragg angle for the reflection θ. At a larger θ, larger ω tilts are required to attain glancing angle conditions. This larger range in ω is visible when the (111) family is compared with the (400).

X-Ray Diffraction and Fluorescence from Rough Surfaces

The Gaussian distribution is commonly used to describe the distribution of excursions (Ref 9, 10). It is also convenient to use for x-ray problems. With this distribution, the density of excursions from the mean surface plane between locations Z_1 and Z_1 + dZ_1 is

$$p(Z) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[\frac{-Z_1^2}{2\sigma^2}\right] \quad \textbf{(Eq 10)}$$

where σ is the standard deviation of the excursions.

The area fraction of sample at a distance Z_1 from the mean plane is

$$A_f = \frac{1}{2} \operatorname{erfc} \frac{Z_1}{\sqrt{2}\,\sigma} \quad \textbf{(Eq 11)}$$

This is the well-known error function complement shown at the right of Fig. 3. It is 0.5 at the mean plane, 0 for large positive excursions beyond about +2σ, and 1 for those deeper than -2σ into the sample.

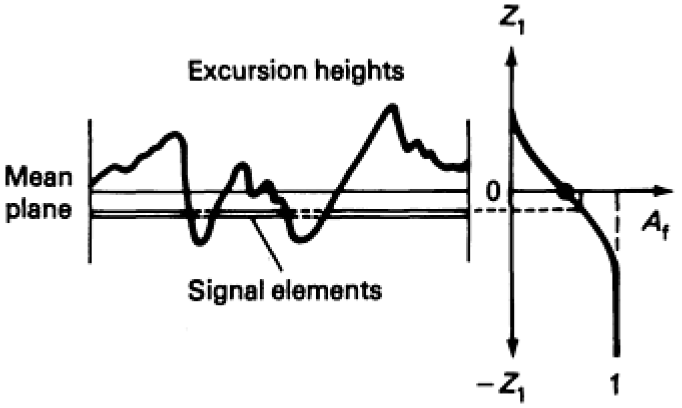


Fig. 3 Location of signal-producing elements about the mean plane of a surface with a Gaussian distribution of asperities. Area fraction, A_f, of occupied sampling plane is shown to right.

X-ray diffraction from polycrystalline samples requires that crystallites be oriented to satisfy Bragg's law. For polycrystalline samples, this usually turns out to be a small fraction of the sampling plane, unless one is using an oriented single crystal. Consequently, the statistical signal fluctuation from XRD can be large for material near the mean plane, unless the material has very small grains. This is not the case for x-ray fluorescence analysis, where all atoms near the surface have a finite probability to contribute a measurable signal.

With a real surface, the signal-producing elements are likely to be correlated with absorbing elements in either the entrance path or along an exit path out of the sample. This surface roughness problem has been treated using numerical calculations based on a Gaussian distribution with correlation (Ref 4). There is no simple analytical answer to this problem. Some general conclusions allow one to establish conditions where surface roughness calculations become unimportant. The x-ray theory contains a correlation parameter, τ_c, which indicates how quickly a surface excursion loses correlation with increasing distance, τ, from a neighboring point. A large τ_c causes neighboring points along the surface to look alike, whereas a small τ_c causes nearby points to be unrelated.

An estimate for the θ value, giving the maximum integrated intensity correction for roughness, can be obtained from

$$\theta_0 = \tan^{-1} \frac{2.4\sigma}{\tau_c} \quad \textbf{(Eq 12)}$$

Likewise, if one is not to exceed a maximum intensity correction of 15%, the following condition should be satisfied:

$$2.4\sigma \leq \mu^{-1} \quad \textbf{(Eq 13)}$$

The term 2.4σ can be related directly to the full width at the half maximum of the Gaussian probability distribution. A 15% correction would represent routine work. For precision work, the maximum σ given by Eq 13 should be decreased by one-half.

Equation 12 describes the condition of having the incident and signal beams oriented parallel to the mean slope of the surface. This shifts the correction to low angles when the correlation parameter τ_C becomes large, relative to σ. Both Eq 12 and 13 treat a sample having a statistically homogeneous distribution of signal elements. When strain gradients that are present near the surface are as large or smaller than 2.4σ, correlation between the signal elements and the exit and entry paths should be considered for a quantitative treatment of the data.

The roughness correction has been shown to go to zero at the extremes of $\theta \rightarrow 90°$ and $\theta \rightarrow 0°$ under symmetrical conditions (Ref 3). At these limits, the paths are either completely correlated or are uncorrelated. The statistical model previously cited (Ref 4) does not treat the case when an incident or signal ray is likely to see more than one asperity.

Experience with the examination of rough surfaces using x-rays is very limited. At this time, x-rays do not appear to give results as good as those obtained using profilometer data in order to determine the statistical parameters that describe real surfaces. At low glancing angles, either with the incident or the signal beam, one can lose quantitative accuracy of intensity data when Z_0 approaches the mean surface excursion distance. However, this need not be a problem for either qualitative chemical or phase identification. Routine approaches for identification problems are described in Ref 8.

Near-Surface Gradients

Equation 2 defines the reduced intensity in terms of a thin slab element having an irradiated area $A_e = A_0/\sin(\theta + \omega)$. For a single-phase homogeneous material free of strain, all elements can be treated as identical and either summed or integrated to obtain the full signal. This led to the integral result for the effective volume, Eq 9. If the surface has been disturbed mechanically, thermally, or chemically, and this produces a change in the spacing of diffracting planes that extends over those probe distances already discussed, the Bragg peaks become distributed over a range of 2θ angles. Peaks are no longer superimposed at the same position and can be treated as differential behavior. Both integral and differential gradients have been described in terms of a continuous distribution function, $H_i(Z)$, which describes the gradient of i with distance Z below the surface (Ref 11, 12). The subscript designates the kind of distribution; that is, residual strain, atom type, a particular phase, texture, grain size, or other structural disturbances distributed along a gradient. These can be defined as integral or differential gradients with respect to the signal elements, depending on whether the intensity is unshifted or shifted in 2θ.

A feature that exhibits a differential behavior displays a change in the ***d***-spacing with position along the gradient zone. Under these conditions, the diffraction angle 2θ between the incident and signals beams changes according to Bragg's law:

$$n\lambda = 2d(Z) \sin \theta(Z) \quad \textbf{(Eq 14)}$$

A change in d causes a distribution of intensity over a range in 2θ angles. The most often leads to an asymmetrical broadening of the Bragg peak, which can be produced by chemical or residual strain gradients. The intensity of the diffracted signal is again proportional to the effective volume or the area of sampling plane multiplied by its thickness, which is given by (Ref 13, 14):

$$\left(\frac{\mathrm{d}d}{\mathrm{d}Z}\right)^{-1} [\exp(-\mu Z S)]\mathrm{d}d \quad \textbf{(Eq 15)}$$

This depends on the d-spacing gradient, $\mathrm{d}d/\mathrm{d}Z$, over each element and leads to a redistribution of the Bragg intensity, along with 2θ axis or differential behavior. Equation 15 indicates that a small d-spacing gradient tends to give higher intensity.

A fluorescence signal "*i*" of a fixed wavelength at each position *Z* gives integral behavior over a gradient zone. When analyzed by a crystal or energy-dispersive detector, one finds that signals at different depths with a common λ superimpose. Therefore, the total signal is obtained by summing all measurable signals to some effective distance below the surface. Here, the signal is not dispersed at each depth -*Z*. Integral behavior is expected from a gradient in the volume fraction of crystal phases having fixed compositions. Here, the diffraction signals accumulate from the entire zone at fixed 2θ positions, giving relatively sharp peaks. The positions are simply determined from Bragg's law and the published lattice parameters and structures (Ref 8) of the various phases along a zone.

Texture can also be treated as an integral quantity because fixed lattice parameters produce an accumulation of diffracted signals from all parts of a zone. If a texture gradient exists over the probe distance considered, the variation in relative integrated intensity from one Bragg peak to another will vary, depending on the pole density gradient of the diffracting planes and the probe depth. A severe thermal-mechanical disturbance at the surface could produce crystallite reorientation and a measurable texture gradient. This problem has not been examined quantitatively.

A grain-size gradient in the range that produces x-ray diffraction line broadening (<100 nm) could give different line widths, again, depending on the gradient, probe depth, and instrument resolution. Because each diffraction peak occurs at the same 2θ position from different portions of the zone, it is considered to be integral behavior. A similar argument can be given for nonuniform strain, but without a uniform strain component. This might be produced by a gradient in the dislocation density. At high dislocation densities, the diffraction peaks become broadened at fixed 2θ positions. This leads to an integral behavior, with either line width or shape changing with probe depth.

There are normally inherent differences in our ability to accurately detect changes in differential and integral measurements. Conventional x-ray diffraction sources are typically of low intensity and require long counting times to attain statistical accuracies better than 5% when polycrystalline samples are used. Integral behavior is examined largely from changes in relative intensity measurements as they vary with different probe depths. Differential behavior further depends on changes in the intensity distribution with 2θ which is usually very sensitive to the gradient profile. The accuracy of the 2θ scale can easily be 1 in 10,000, making an XRD measurement of differential behavior highly accurate. The following sections provide examples of both integral and differential behavior using XRD.

Integral Gradient

The examination of phase gradients along a wear track of a partially stabilized zirconia disk by XRD provides an example of integral behavior. Three distinct crystal phases can be present in this material: cubic, tetragonal, and monoclinic. Linear, exponential, and stepped distributions have been considered (Ref 11, 12). The exponential distribution can be written as (Ref 11):

$$H_i(Z) = H_{0i} + H_{si}e^{-Z/b_i} \quad \textbf{(Eq 16)}$$

The various terms are illustrated and defined in the caption of Fig. 4. For each position *Z*, it is required that

$$\Sigma H_i (Z) = 1 \quad \textbf{(Eq 17)}$$

where all phases *i* are considered. The effective volume for each phase is obtained by integration

$$V(ij\lambda\omega) = \frac{A_0}{\sin(\theta + \omega)} \int_{-\infty}^{0} H_i(Z) \exp(-\langle\mu\rangle SZ)\, dZ \quad \textbf{(Eq 18)}$$

with

$$S = \frac{1}{\sin(\theta + \omega)} + \frac{1}{\sin(\theta - \omega)} \quad \textbf{(Eq 19)}$$

and the subscript *i* refers to the Miller indices (*hkl*) reflecting at an angle θ. The linear absorption coefficient averaged over all phases is $\langle\mu\rangle$. For the exponential distribution given by Eq 11, this simplified to

$$V(ij\lambda\omega) = \frac{A_0 H_{0i}}{S\langle\mu\rangle \sin(\theta+\omega)} \left[1 + \frac{b_i \langle\mu\rangle S(H_{si}/H_{0i})}{1 + b_i \langle\mu\rangle S}\right] \quad \text{(Eq 20)}$$

for a flat sample.

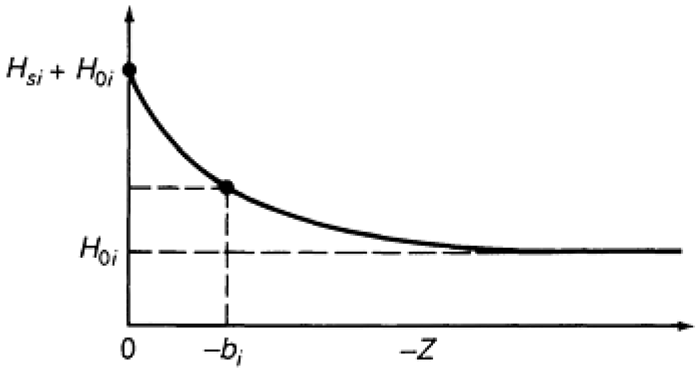


Fig. 4 Experimental near-surface area fraction distribution showing: undisturbed substrate limit, H_{0i}; surface area fraction, H_{si}, and location e^{-1} point in terms of $-b_i$

The quantity *S* is the path length factor given in Eq 8 with $\mu_s = \mu_i$. *S* becomes very large as $\omega \rightarrow \pm\theta$. Two cases are considered below.

In the first case, the beam samples only a small fraction of the exponential distribution. This condition requires that $(b_i\langle\mu\rangle S) \gg 1$. A large value for any one of these terms could limit the overall view of the gradient. Therefore, it is reasonable that the effective volume for phase *i* should be

$$V(ij\lambda\omega) = \frac{A_0(H_{0i} + H_{si})}{S\langle\mu\rangle \sin(\theta+\omega)} \quad \text{(Eq 21)}$$

That is, only the phase distributions at the surface are seen.

In the second case, if $(b_i\langle\mu\rangle S) \ll 1$, due to any one or any combination of these terms, the near-surface gradient is too small to be observable. Therefore, the effective volume is determined by the volume fraction in the substrate

$$V(ij\lambda\omega) = \frac{A_0 H_{0i}}{S\langle\mu\rangle \sin(\theta+\omega)} \quad \text{(Eq 22)}$$

In order to evaluate b_i, at least one measurement should be made in the range ($b_i\langle\mu\rangle S \approx 1$, and additional measurements are required to solve for H_{0i} and H_{si}. These are obtained at different values or with different μ values, because of a λ change.

The determination of the phase distributions for a polished PSZ dish after a wear test was conducted for both the worn track and an unworn region (Ref 11). The wear testing was conducted at an ambient temperature of 204 °C (400 °F), with an atmosphere of prepurified nitrogen. A 6.4 mm (0.25 in.) diameter pin was loaded to 12.5 N (1.27 kgf) against a disk

traveling at 0.3 m/s (1 ft/s) at the point of contact. The total time of testing was 61 h. XRD patterns from both the worn track and the back side of the disk were obtained with synchrotron radiation of wavelength λ= 0.24797 nm. Two paths were obtained using a typical symmetrical arrangement (ω = 0°) and an asymmetrical arrangement having the sample normal tilted toward the incident beam, giving a 6° angle with the diffracted beam. The XRD patterns include the (111) and ($11\bar{1}$) reflections of the monoclinic phase [M(111), M($11\bar{1}$)] and the overlapping (111) reflections of the cubic and tetragonal phases [C(111)] and [T(111)], respectively.

Figures 5 and 6 show the general features of the symmetrical and asymmetrical diffraction patterns. The peak separation was carried out using a Pearson VII least squares fitting procedure. Although wear testing introduces a detectable amount of the tetragonal phase, the combined intensity of C(111) and T(111) are used, along with M(111) and M($11\bar{1}$). These will be used later to determine the phase distributions of the combined cubic and tetragonal phases. The combined C(111) and T(111) reflections are denoted by CT(111).

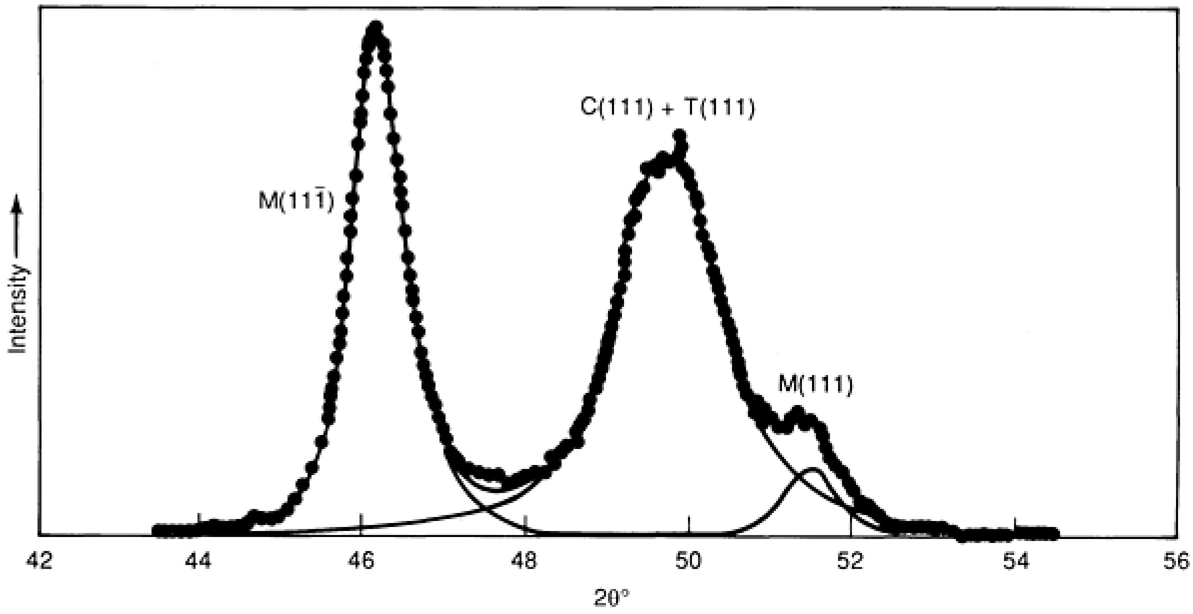


Fig. 5 X-ray diffraction data points for unworn side of a PSZ disk obtained using symmetrical optics and λ= 0.24797 nm. Individual profiles are shown as solid lines. Source: Ref 11

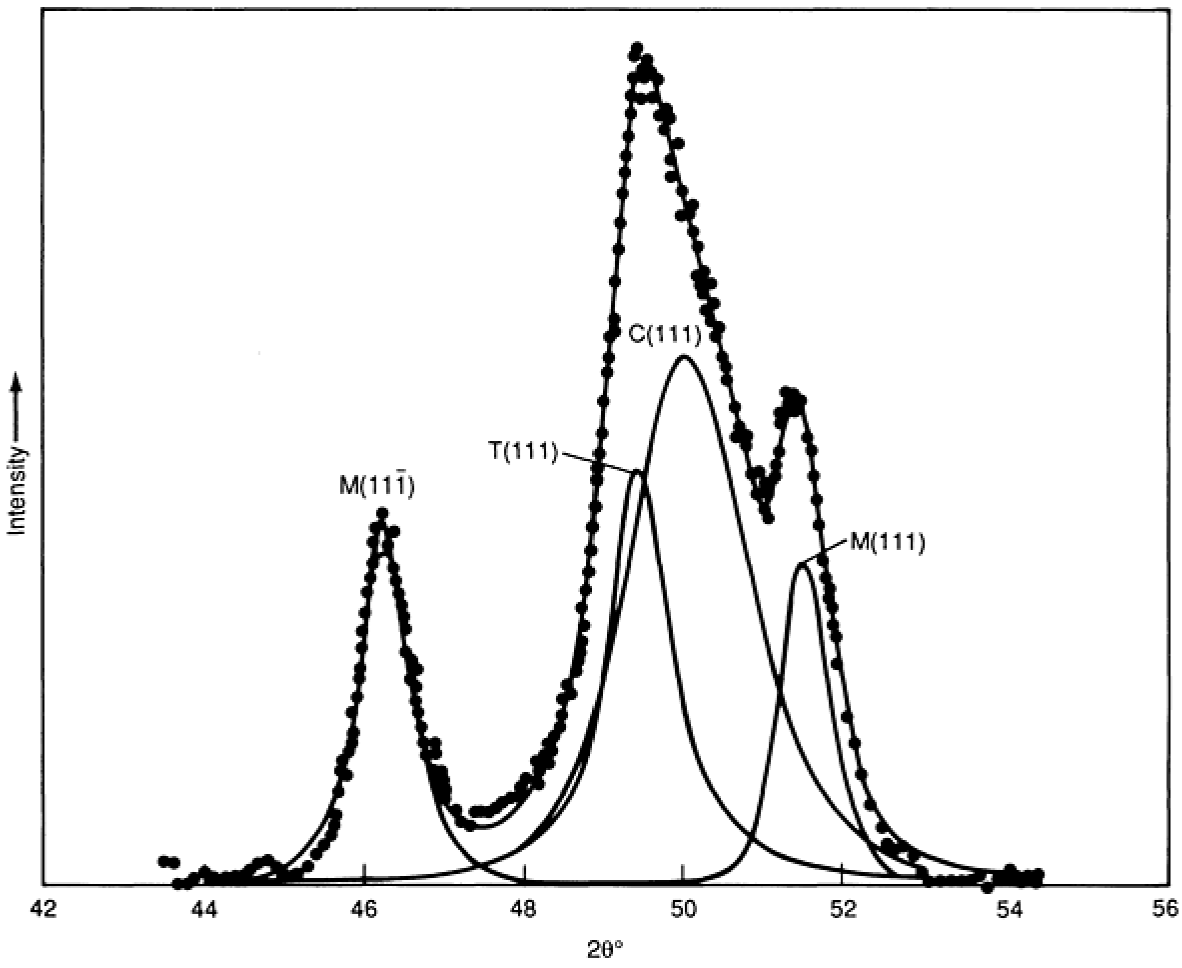


Fig. 6 X-ray diffraction data points for worn side of PSZ disk obtained with asymmetrical optics and λ= 0.24797 nm. Individual profiles are shown as solid lines. Source: Ref 11

A constant volume fraction of the undisturbed bulk region was obtained from the back side of the PSZ disk with CuK$\alpha 1$ radiation and symmetrical diffraction geometry. The linear absorption coefficient of PSZ for CuK$\alpha 1$ radiation is $\approx \frac{1}{4}$ of that for the λobtained with synchrotron radiation. This combination results in deeper penetration of the x-ray beam, which gives an XRD pattern that better represents the phase distribution in the bulk material. These intensities gave an upper limit of 0.445 for the volume fraction of cubic and tetragonal phases (H_{oct}). Profiles for both the worn track and the polished back side of the disk are shown in Fig. 7.

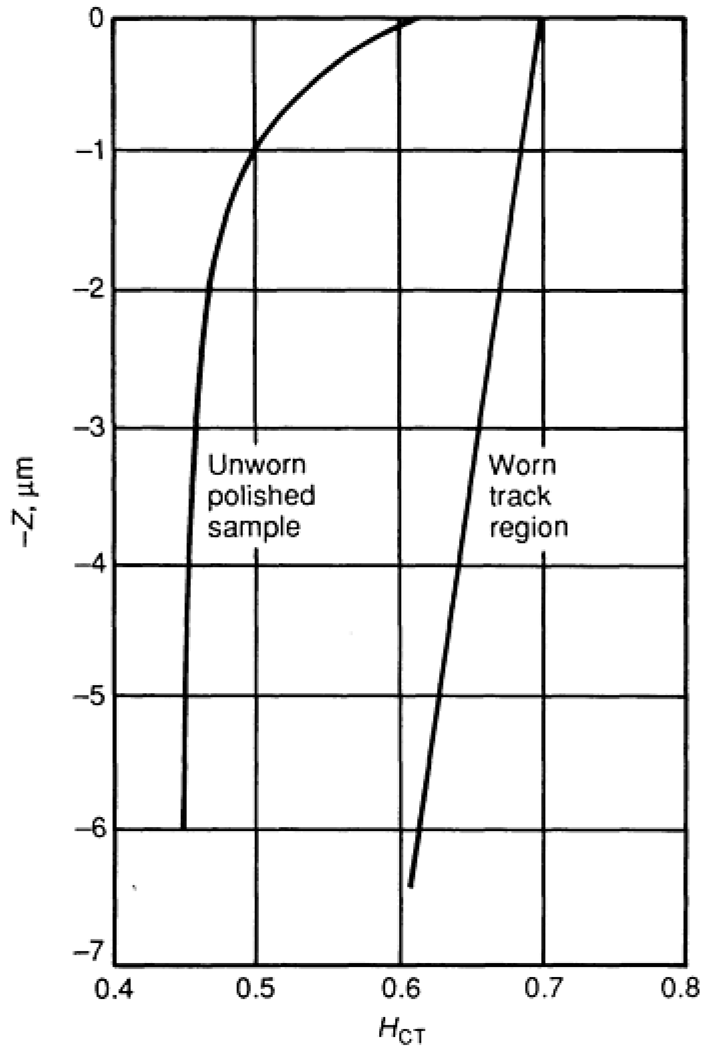


Fig. 7 Distribution of high-temperature cubic and tetragonal phases for unworn polished sample and worn track region. Source: Ref 11

Different probe lengths are obtained by adjusting $S<\mu>$. The ratio of the reduced intensity of one phase to another is directly related to the effective volume ratio

$$R_{ii'jj'\lambda\omega} = \frac{\sum_j I'_{ij\lambda\omega}}{\sum_{j'} I'_{i'j'\lambda\omega}} = \frac{\sum_j V_{ij\lambda\omega}}{\sum_{j'} V_{i'j'\lambda\omega}} \quad \textbf{(Eq 23)}$$

The notation is defined as follows: *i* and *i*' designate any two phases in a sample with *n* phases. The summations over *j* and *j*' indicate that more that one peak can be selected from each of the phases. Because *i* and *i*' can be any two of the *n* phases, Eq 23 represents a set containing up to (*n* - 1) independent intensity ratios for a fixed λ and ω. In the multiple-wavelength technique (Ref 12), one normally uses ω= 0° in Eq 23 to get 2 (*n* - 1) intensity ratios in terms of *n* effective volumes. With the integrated intensity ratios measured from the XRD patterns, the unknown parameters contained in the effective volumes can be determined because μ varies with λ. Similarly, with multiple-beam paths, one can use a fixed λ and several tilt angles, ω, to obtain the intensity ratios in terms of *n* effective volumes. The unknown parameters, located in the effective volumes, can be obtained from the measured integrated intensities and other data that are normally available.

Differential Gradient

Differential behavior is observed when the interplanar spacing, *d*, changes with depth below the free surface and intensity becomes redistributed along the 2θ axis. The *d*-spacing change can be relatable to composition changes resulting from small differences in atomic size or it can be due to residual strain from thermal and/or mechanical interactions at the free surface. Composition gradients have been most extensively studied using XRD (Ref 13, 14).

In the preceding example of integral behavior, a partially stabilized zirconia was used to show a phase gradient. For the present discussion, it is instructive to continue with zirconia, but to use one that is fully stabilized (FSZ) in the cubic form. Two samples were examined (Ref 15). One was polished to give σ= 0.027 μm, and a second was severely ground to give σ= 6.2 μm and a correlation distance (τ_c) of 81 μm. For the polished sample, the beam penetration distance was always much larger than σ, whereas for the ground sample, σ was always at least a factor of 3 larger.

The (111) peak profiles of the polished and the ground FSZ samples were obtained using Cuk$\alpha 1$ radiation and symmetrical diffraction optics. The optics. The CuK$\alpha 1$ component was obtained using a diffracted beam quartz monochromator and a fine-focus Cu tube. Results from the polished and ground samples are shown in Fig. 8. Profiles extend asymmetrically toward the low-angle side, because of a state of compression near the surface. The (111) profiles of the polished and the ground samples were also measured, using synchrotron radiation of 0.24794 nm with both symmetrical and asymmetrical optics, at the Oak Ridge National Laboratory beamline (National Synchrotron Light Source), before annealing. The polished and ground sample profiles measured with symmetric diffraction optics are shown in Fig. 8(a) through (d). Intensity bands are observed from ground and polished samples using two radiations. The longer synchrotron wavelength is less penetrating and better emphasizes the distribution near the surface zone.

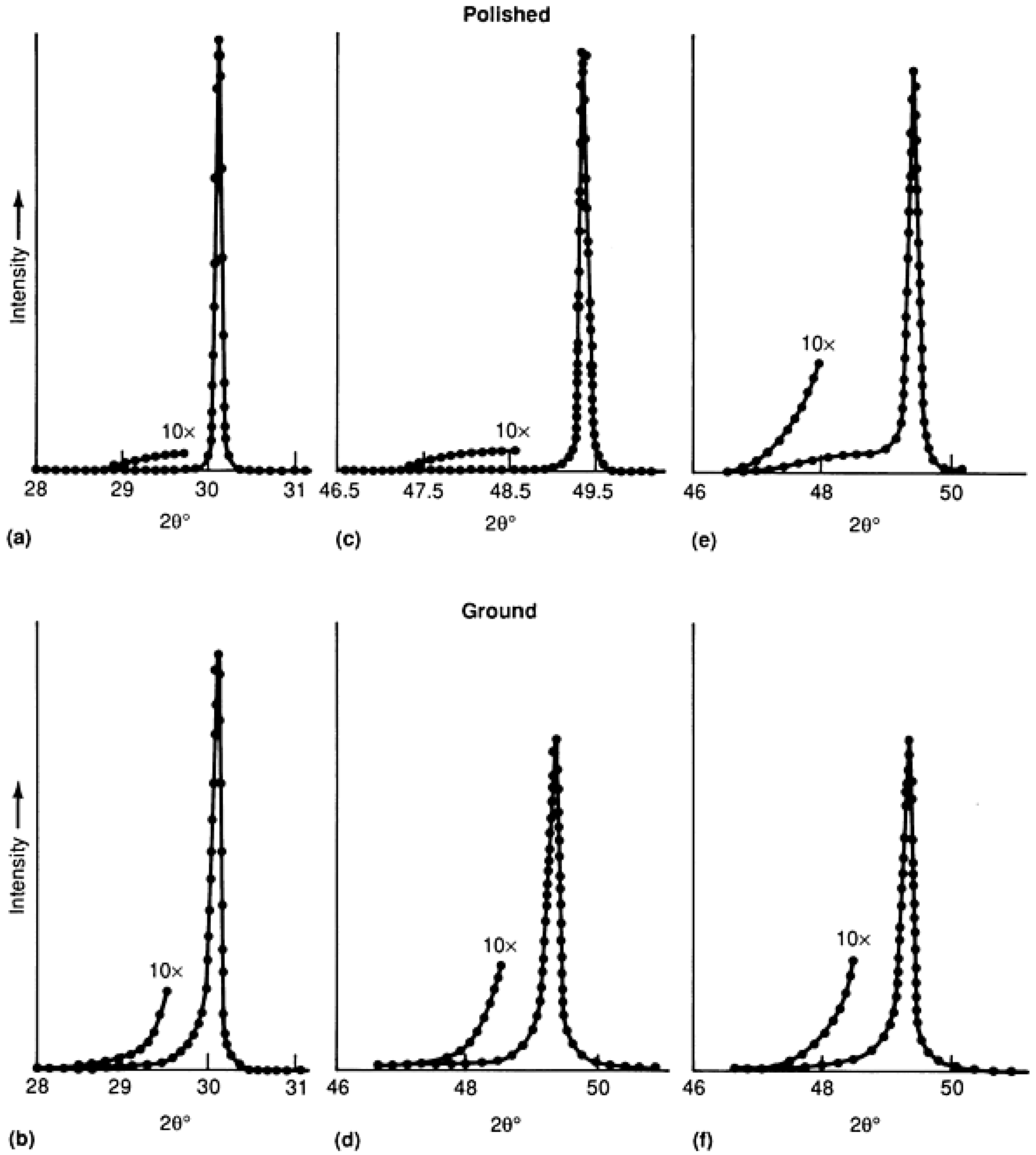


Fig. 8 (111) intensity bands from polished and ground samples of partially stabilized zirconia showing experimental data points and computer simulations (solid lines) for and symmetrical optics with CuKα radiation (λ - 0.15406 nm), (a) and (b); synchrotron radiation with λ= 0.24797 nm, (c) and (d); and 3° asymmetric optics with λ= 0.24797 nm, (e) and (f)

To further emphasize the near-surface regions, (111) profiles of both samples were measured with asymmetrical diffraction optics, as shown in Fig. 8(e) and 8(f). Asymmetric diffraction optics were obtained by tilting the sample normal toward the diffracted beam by an amount such that the incident angle is 3°. The small incident angle results in less penetration, and, as expected, influences the peak profile for the polished sample. The peak profile of the polished sample obtained with asymmetrical diffraction optics enhances the low-angle side of the pattern compared to that found with the symmetrical optics. For the ground sample, no change was observed between the profiles obtained with symmetrical and asymmetrical optics. This result will be discussed later.

After both samples were subsequently annealed at 1200 °C (2190 °F) for 1 h and furnace cooled, the profiles were indistinguishable. However, considerable sharpening was observed. To check these findings, the samples were repolished

using alumina powder. No peak shift was observed, but the profile was extended asymmetrically toward the low-angle side. This reconfirms the conclusion that the polishing process introduces a low-level intensity band, and the main peak remains unshifted. The peak shift between the profiles of the ground and polished samples before annealing is real, and is due to the extended range of the strain gradient in the severely ground sample.

With depth measured from a load surface asperity, the sample can be treated as a system of thin curved layers located at various depths. The *d*-spacing of each layer is treated as a constant, and the corresponding intensity is calculated using a theory that includes roughness. Figure 9 illustrates two strain profiles obtained from two radiations for the severely ground sample. The polished sample is essentially flat, with the layers parallel to the sample surface, whereas the ground sample is rough, and consists of layers of constant strain that follow the same fluctuation as the surface profile. The diffraction theory treats the intensity from each layer under both the symmetrical and asymmetrical diffraction conditions. The strain profiles for the polished sample are shown in Fig. 10. Because the *d*-spacing changes continuously from one layer to another, the intensities are spread over a 2θ range according to Bragg's law.

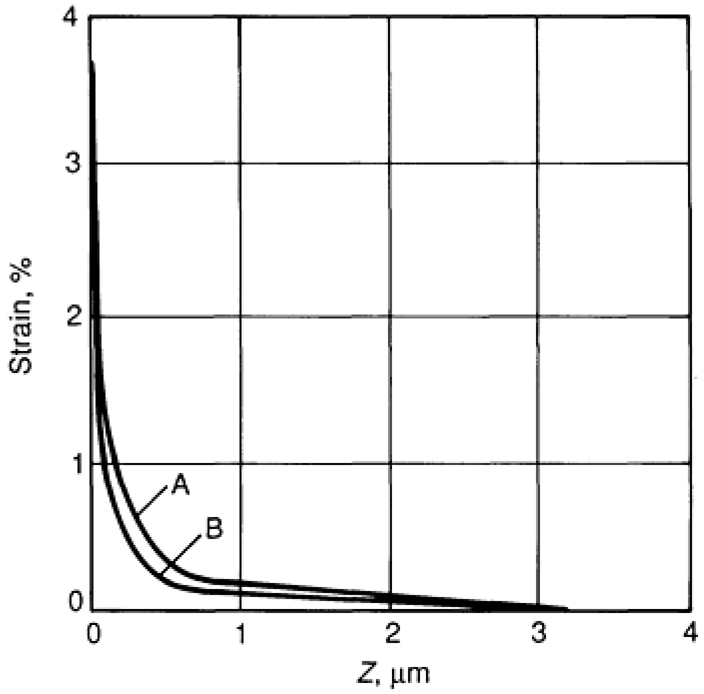


Fig. 9 Depth profiles of strain for ground sample obtained with symmetrical optics. A, CuKα radiation; B, λ = 0.24797 nm (synchrotron radiation). Source: Ref 15

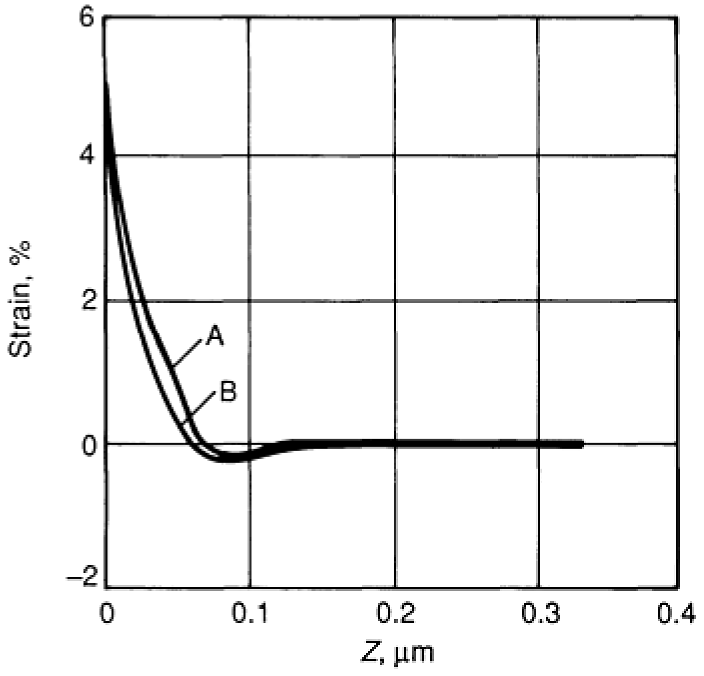


Fig. 10 Depth profiles of strain for polished sample obtained under the following conditions: A, symmetrical optics using CuK$\alpha 1$ (λ = 0.15406 nm and λ= 0.24797 nm) (synchrotron radiation); B, asymmetrical optics with λ= 0.24797 nm (synchrotron radiation). Source: Ref 15

Computer-simulated intensity bands are compared with the fitted experimental data in Fig. 8(a) through (f). Both the measured and simulated (solid lines) profiles are expanded vertically by 10 times at the low-angle side to better show the final fit. The redistribution of intensity, in this case, is weak and could be overlooked in routine XRD. It could be completely missed in complex powder patterns having many overlapping diffraction peaks.

A reexamination of Fig. 9 and 10 shows that the maximum compressive strain near the surface is large, giving $\sim$4% for the ground surface and $\sim$5% for the polished. Although polishing gives the highest strain, the overall zone thickness is about $\frac{1}{10}$ the size found after severe grinding.

Analysis of Debris Particles

At one instant of time during a wear test, a small fraction of the asperities of a surface become separated as debris particles. It would be very instructive to be able to focus only on the characteristics of these regions, that is, to determine the crystal phases, dislocation density, and a statistical description of their arrangement. However, these features are lost in an examination of the overall near-surface region because of their small volume fraction.

An alternative is to device efficient methods of collecting the debris particles that break away from these special asperities. An analysis of debris particles could be useful in establishing wear mechanisms. This collection need not require the dismantling of a machine, but rather, a concentrate of wear fragments could be collected at filter locations. Ideally, one would like to look back and relate the structural parameters of debris to the conditions present in the asperities prior to breakaway, and to correlate these findings with machine operational conditions. The previous discussion of substrate structure relates to wear conditioning processes. Both substrate and debris should be examined in a comparative approach.

For material that contain more than one crystal phase, or a material that can undergo further phase transformation, the first question one might ask is: What phases are present? By knowing the stability of these phases from either the equilibrium diagrams or the temperature range of metastable, phases, one can relate back to thermal and chemical conditions prior to the breakaway point. This is particularly true if the particles are small and have a high surface/volume ratio, allowing

rapid cooling from the local temperature at an asperity during separation. An example of a debris analysis is given later for zirconia.

X-ray diffraction profile analysis has been used to examine highly deformed metals since the 1950s. More recently, this has been described in Ref 6 and 7. The results provide a mean subgrain size and root mean square strain as a function of distance and crystal direction. Early investigations were carried out on "cold-work" filings. In other words, powder samples were obtained by filing a solid sample whose average temperature was close to room temperature. Any local heating associated with the breakaway process was overlooked in what was considered to be a cold-working process. The most reliable data of this type have been reexamined (Ref 15, 16). The process of producing filings can be treated as one form of wear process. A major concern of line-shape techniques is the need to examine the full profile as pairs of first- and second-order Bragg peaks. When profiles overlap for one reason or another, one must make procedural compromises.

Recent advances allow a least squares fitting procedure to be used on overlapping data sets (Ref 16). It gives the mean particle size and two strain parameters, $<\epsilon^2_{1U}>$, $<\epsilon^2_{1D}>$. The mean square strain varies according to

$$<\epsilon^2_{n}> = <\epsilon^2_{1U}> + <\epsilon^2_{1D}>/n \quad \textbf{(Eq 24)}$$

where n gives distance between two points in the crystal in units of the average d-spacing, d, for the first order of a set of reflections ($h_1k_1l_1$) and ($h_2k_2l_2$). Strain is determined along columns perpendicular to the reflecting planes. Knowing the two strain parameters given above allows $<\epsilon^2_{n}>$ to be evaluated. A pair of fitted profiles obtained from zirconia debris are shown in Fig. 11 (Ref 17).

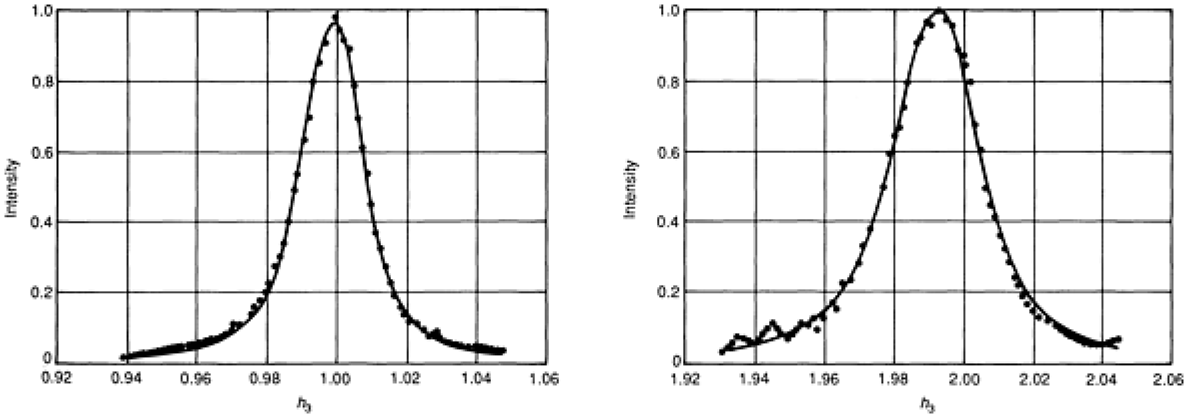


Fig. 11 Least squares fitting of (200) (left) and (400) (right) profiles using procedures described in Ref 18

Equation 24 presents the results in a purely statistical form, and does not provide a basic connection with dislocation theory. It has been shown that the two strain parameters noted above, as well as $<L>$, the average subgrain size, determine the dislocation density according to (Ref 18, 19):

$$n_d = 122\,\frac{d}{b^2}\,\frac{\langle \epsilon^2{}_{1D}\rangle}{L} \quad \textbf{(Eq 25)}$$

where b is the Burger's vector.

This varies directly with $<\epsilon^2_{1D}>$, and inversely with subgrain size, L. A second parameter, obtained from the more-recent line-shape analysis, contains the correlation distance, R_c, for unlike dislocation relative to L. This is given by (Ref 17, 18):

$$\frac{R_c}{L} = 1.10 \exp\left(\frac{0.62L\langle\epsilon^2{}_{1U}\rangle}{d\langle\epsilon^2{}_{1D}\rangle}\right) \quad \textbf{(Eq 26)}$$

Isotropic elasticity theory has been used to obtain these relationships for dislocation density and correlation parameters. In applying Eq 24, one need not assume that the material is isotropic.

Table 1 lists results for both face- and body-centered cubic materials. A common pattern is apparent. For cold-work filings, one finds that

$$1.33 \leq \frac{R_c}{\langle L\rangle} \leq 2.17 \quad \textbf{(Eq 27)}$$

For the most part, the dislocation density is high and in the $10^{11}/cm^2$ range. Dislocations are highly correlated, with R_c close to the average diameter of the subgrain for metallic filings and ZrO_2 debris.

Table 1 Results for body-centered cubic materials, face-centered cubic aluminum, and partially stabilized cubic zirconia

Material	<L>, nm	$n_d \times 10^{-11}/cm^2$	R_c/<L>	$\langle\epsilon^2_{1D}\rangle \times 10^4$	$\langle\epsilon^2_{1U}\rangle \times 10^6$
Cr	28.5	3.67	1.79	2.63	1.47
V	20.0	6.56	2.17	3.46	4.05
Mo	26.0	4.60	1.99	3.30	2.72
W	21.5	7.32	1.66	4.56	3.27
Nb	20.0	1.53	1.33	0.84	0.29
Al	32.0	4.41	1.64	2.32	1.60
ZrO_2	8.6	12.70	1.42	4.84	4.90

Aluminum filings are prepared under liquid nitrogen.

Source: Ref 18

The results given in Table 1 for body- and face-centered cubic material allow a simple explanation to be given for the structure of cold-worked fragments. Filings show a high degree of correlation (small R_c/<L>) and high dislocation densities. This is related to sample conditions, such as the temperature and local conditions at which a high-density dislocation structure is produced. Fragments produced by filing or grinding at room temperature, or even at liquid nitrogen temperature, cannot be "cold." Instead, they are rapidly heated to a high temperature during fracture, and rapidly quenched, because of their high surface-to-volume ratio. Dislocations are mobile for short periods of time over short distances and can cluster.

Table 1 includes results from cubic debris particles obtained from a worn PSZ. The substrate contains three phases, as already discussed. It should be noted that the cubic phase is stable at high temperatures, making it reasonable that only the cubic form should be found in the debris. The dislocation density for zirconia is the highest entry, whereas the correlation distance is close to the mean diameter. One can present the same arguments for zirconia debris as for cold-worked metal filings. Prior to breakaway, material at the asperity has a very high dislocation density. A high separation temperature allows some dislocation movement to form subgrain clusters.

The question of how much debris one must collect for an ideal line-shape analysis needs to be considered. Typically, for quantitative work, one would like to be able to fill a cavity of at least 15 mm (0.6 in.) in diameter to a depth of 2 mm (0.08 in.). The sample thickness should be several times the penetration depths already discussed. Quantitative diffractometer studies require much more sample than what is required for an x-ray powder camera, but also offer the opportunity for more extensive data analysis.

Conclusions and Future Trends

Both the wear-modified phase distribution and the debris indicate that at least a fraction of the surface material has been subjected to high temperatures during wear testing. Zirconia debris has an even higher dislocation density than metal filings, indicating severe deformation before breakaway from the surface. High correlation between dislocations is likely to result from a time-restricted, thermally activated process taking place at an elevated temperature immediately after fracture. It was found that large residual compressive strains build up in the ground surface region of the zirconia substrate. A state of dynamic equilibrium is likely to be present between the production of regions of high defect density-residual strain and relaxations produced by annealing. The annealing processes within near-surface bulk substrate material is likely to offer a spectrum of possibilities.

Data from a ground PSZ sample revealed the formation of a place gradient extending over a distance of several microns. This represents another form of conditioning that could influence subsequent wear testing. Depths from a few tenths to the micron range are typical x-ray probe distances for many commercial materials.

Future research will require a careful selection of samples and radiation. Quantitative diffractometer data will give the greatest amount of information. A part from local surface asperities, the mean surface should be flat. Typically, it also should have surface dimensions ranging from 10 to 30 mm (0.4 to 1.2 in.) to fit into commercial systems. XRD data should be inter-related with XRF or other near-surface analyses revealing chemical changes. Variable penetration depths often allow the condition near a surface to be compared with deeper regions that are relatively unaffected by wear processes.

A detailed analysis of diffraction patterns from concentrated industrial wear debris under different conditions may establish trends that predict malfunctions. Similarities or differences between the wear from a given machine and testers, which are used to simulate machine conditions, could also be examined.

Thin-film attachments are available for most commercial diffractometers, allowing low glancing angles to be attained. This can greatly reduce the beam penetration. These attachments should be considered for examinations of worn surfaces, as well as for thin layers of debris.

Perhaps the most severe restriction in an x-ray analysis using conventional x-ray sources is the limited area of disturbed surface available for examination using routine pin-on-disk testing. One would rather examine square cross sections of at least 15 mm (0.6 in.). Although synchrotron radiation would allow one to examine a small fraction of this size, it is often not readily available. A smaller-sized sample used with conventional sources would force compromises in the data analysis and lessen the opportunity to obtain quantitative results. Any use of x-rays must begin with a consideration of sample size.

References

1. G.H. Vineyard, *Phys. Rev. B: Condens. Matter,* Vol 26, 1982, p 416
2. W.C. McMaster, N. Kerr Del Grande, J.H. Mallett, and J.H. Hubbell, "Compilation of X-ray Cross Sections," Report UCRL-50174, Sec. I, 1970; Sec. II, Rev. I, 1969; Sec. III, 1969; Sec. IV, Lawrence Radiation Laboratory (Livermore), 1969
3. R.J. Harrison and A. Paskin, *Acta Cryst.,* Vol 17, 1964, p 325
4. B. Hwang and C.R. Houska, *J. Appl. Phys.,* Vol 63, 1988, p 5346
5. C.J. Sparks, *Synchrotron Radiation Research,* H. Winick and S. Doniach, Ed., Plenum Publishing, 1980
6. B.E. Warren, *X-Ray Diffraction,* Addison-Wesley, 1969
7. L.H. Schwartz and J.B. Cohen, *Diffraction from Materials,* Springer-Verlag, 1987
8. B.D. Cullity, *Elements of X-Ray Diffraction,* Addison-Wesley, 1978
9. B. Hwang, "Near Surface Structure of Ceramic Components," Ph. D. thesis, Virginia Polytechnic Institute and State University, May 1987
10. T.R. Thomas, *Rough Surfaces,* Longman, London, 1982
11. B. Hwang, C.R. Houska, G.E. Ice, and A. Habenschuss, *Adv. Ceram. Mater.,* Vol 3, 1988, p 189 V
12. R.C. Garvie, R.H.K. Hannink, and N.V. Swain, *J. Mater. Sci. Lett.,* Vol 1, 1982, p 437
13. C.R. Houska, *J. Appl. Phys.,* Vol 41, 1970, p 69
14. C.R. Houska, *Treatise on Materials Science,* Vol 19A, H. Herman, Ed., Academic Press Inc., 1980

15. B. Hwang, C.R. Houska, G.E. Ice, and A. Habenschuss, *J. Appl. Phys.,* Vol 63, 1988, p 5351
16. C.R. Houska, *J. Appl. Phys.,* Vol 52, 1981, p 748
17. S. Rao and C.R. Houska, *Acta Cryst.,* Vol A42, 1986, p 14
18. S. Rao and C.R. Houska, *Matter. Res. Soc. Symp. Proc.,* Vol 138, 1989, p 93
19. S. Rao and C.R. Houska, *Acta Cryst.,* Vol A44, 1988, p 1021

Basic Tribological Parameters

Horst Czichos, BAM (Germany)

Introduction

TRIBOLOGICAL PARAMETERS are characteristics of mechanical systems with "interacting surfaces in relative motion," including the initiation of motion. The tribological processes of interacting surfaces have a dual character. They are on one hand necessary for the functional performance of "tribosystems" or "tribocomponents" (see the "Glossary of Terms" in this Volume), but are on the other hand inevitably connected with friction and wear. In engineering applications, the functional purpose of trobosystems can be broadly classified into the following categories (Ref 1):

- The guidance, transmission, coupling, control, stop, and annihilation of motion, force, mechanical energy, and power (bearings, joints, gears, clutches, cams and tappets, bolts and nuts, fasteners, and brakes)
- The transportation and control of flow of matter (pipelines, wheel/rail, tire/road, valves, and seals)
- The forming, machining, and tearing of materials (drawing, pressing, cutting, shaping, quarrying, and dredging)
- The generation and transmission of information (printing heads and magnetic recording interfaces)

The diagnosis of friction and wear data of such tribosystems or corresponding laboratory test configurations and test specimens requires special attention because numerous characteristics, parameters, and factors must be taken into consideration. This is due to the fact that friction and wear are not intrinsic materials properties, but must be related to the entire system of interacting components, namely materials pairs and interfacial lubricants. This is obvious from a comparison between the test conditions to obtain strength data or friction and wear data (Fig. 1).

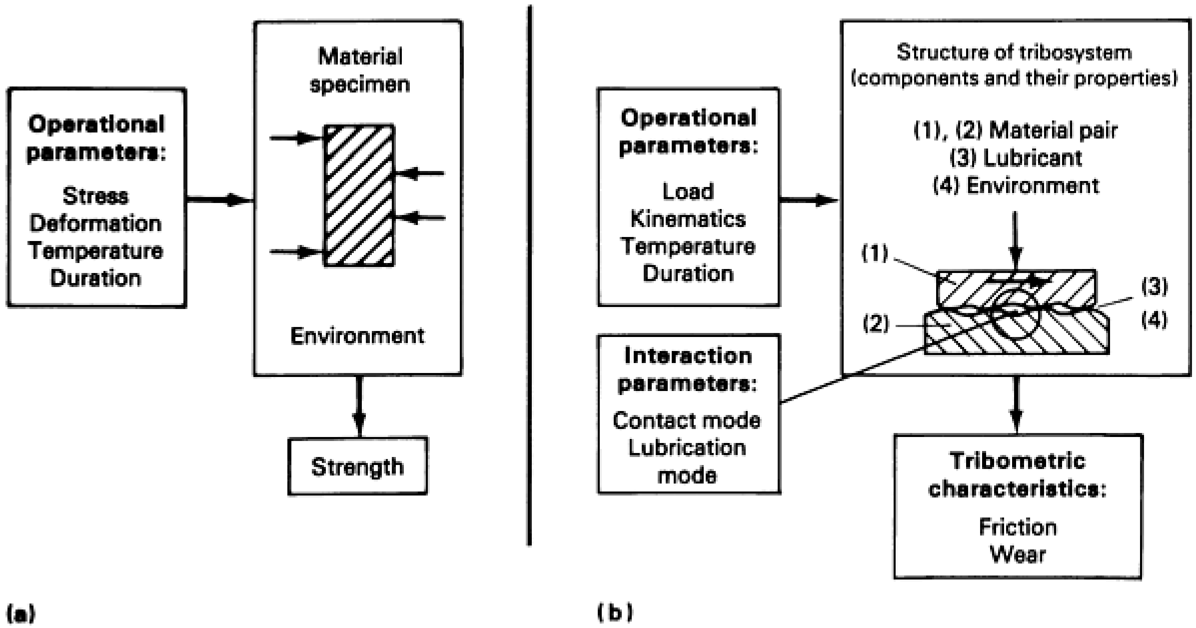


Fig. 1 Characteristics and parameters of (a) strength tests and (b) friction and wear tests

In strength tests (Fig. 1a), the deformation or fracture resistance of a material specimen in a given environment is determined under the action of a certain stress mode, such as tension, compression, shear, bending, or torsion. The

resulting strength data (in terms of force per cross section, or energy) are considered as intrinsic materials properties depending basically on the following groups of parameters:

- *Materials parameters,* such as composition, microstructure, and specimen geometry
- *Operational parameters,* such as stress type, load, deformation velocity, and temperature

In a friction or wear test (Fig. 1b), the resistance against motion (friction) or the resistance against surface damage (wear) of a material/material pair (dry system) or a material/lubricant/material combination (lubricated system) in a given environment is determined under the action of a certain type of motion, such as sliding or rolling. The resulting tribometric characteristics--in particular the friction or wear data--must understood as tribological systems characteristics associated with the following group of parameters:

- *Structural parameters,* which characterize the components (materials, lubricant, and environment) involved in the friction and wear process and their physical, chemical, and technological properties
- *Operational parameters,* that is, the loading, kinematic, and temperature conditions and their functional duration
- *Interaction parameters,* which characterize, in particular, the action of the operating parameters on the structural components of the tribological system and define its contact and lubrication modes

Structural Parameters

The analysis of structural parameters must identify first the components involved in a given friction and wear problem. Figure 2 shows typical examples of tribosystems subject to friction and wear together with corresponding simplified test configurations and their elementary structure. This figure illustrates that in any friction and wear situation, four tribocomponents are involved (Ref 2):

- Triboelement (1)
- Triboelement (2)
- Interfacial element (3), for example, lubricant or dust particles
- Environmental medium (4), for example, air or corrosive atmosphere

Table 1 lists examples of the tribocomponents that make up various tribosystems.

Table 1 Structural components of common tribosystems

Tribosystem	Triboelement (1)	Triboelement (2)	Interfacial element (3)	Environmental medium (4)	Type of systems structure
Gear box	Gear 1	Gear 2	Gear oil	Air	Closed
Wheel/rail	Wheel	Rail	Moisture	Air	Open
Sliding guide	Slider	Support	Grease	Air	Closed
Bearing	Bushing	Shaft	Lubricant	Oil mist	Closed
Dredge	Shovel	Soil	. . .	Dust	Open
Milling system	Milling wheel	Milling jaw	Minerals	Air	Open

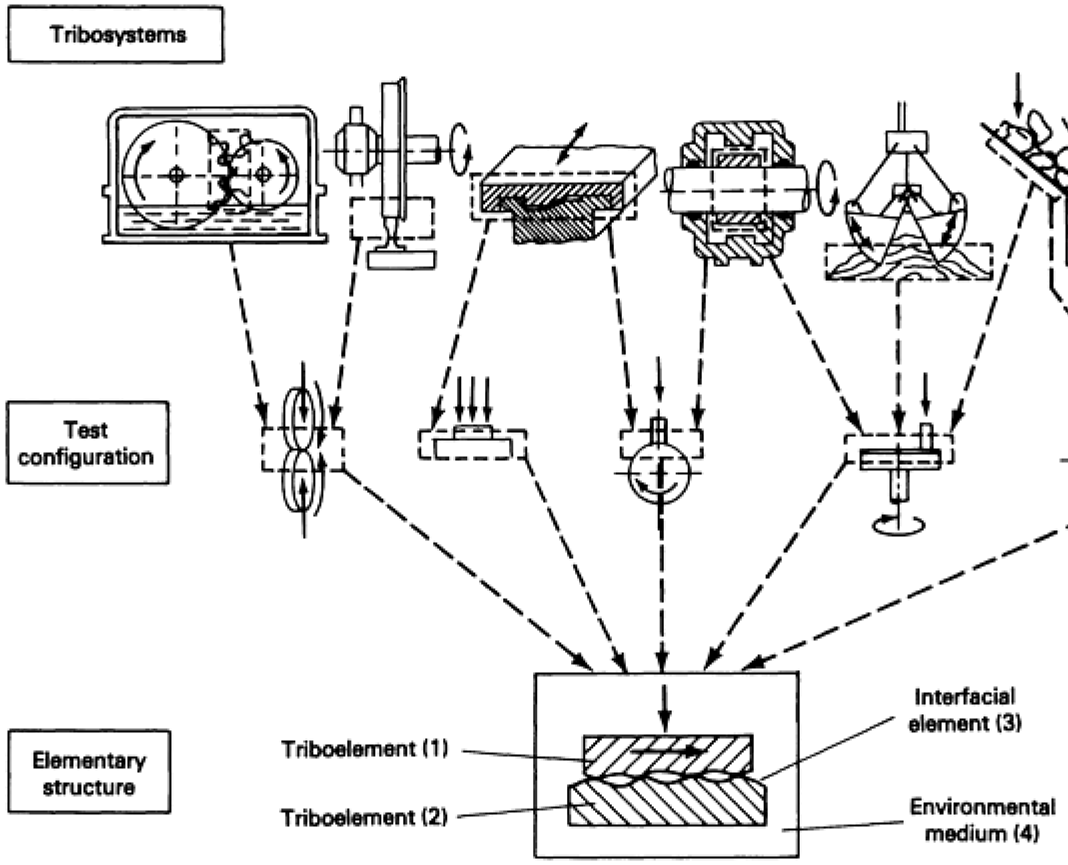


Fig. 2 Examples of engineering tribosystems, test configurations, and their elementary structure

In analyzing the structure of tribosystems, a distinction can be made between "closed systems," in which all components are continuously involved in the friction and wear process, and "open systems," in which a materials flow in and out of the system occurs.

The friction and wear data of tribosystems depend on various properties of their structural components (tribocomponents). Structural parameters of closed tribosystems can be classified in most cases into two groups.

Group A consists of triboelements (1) and (2) and involves:

- *Chemical parameters* such as volume composition and surface composition
- *Physical parameters* such as thermal conductivity
- *Mechanical parameters* such as elastic modulus, hardness, and fracture toughness
- *Geometric parameters* such as geometry dimensions, and surface topography
- *Microstructural parameters* such as grain size, dislocation density, and stacking fault energy

Group B consists of interfacial (fluid) element (3) and environmental (gaseous) medium (4) and involves:

- *Chemical parameters* such as composition, additive content, acidity, and humidity
- *Physical parameters* such as density, thermal conductivity, and flash and fire point
- *Mechanical parameters* such as viscosity, and viscosity-temperature and viscosity-pressure characteristics

For open systems, for example, manufacturing systems such as machining and molding, or quarrying and dredging systems, the structural parameters characterizing the materials flow in and out of the system are often difficult to specify.

In addition to the structural elements necessary to fulfill the functional purpose of the tribosystem, detrimental elements such as dirt, dust, and moisture may also be present and must be recognized in the analysis of structural parameters. To assist in the compilation of the various parameters relevant to a given friction and wear problem, a data sheet of basic tribological parameters is described later in this article (see the section entitled "Data Sheet of Basic Tribological Parameters").

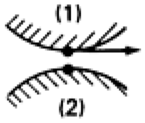
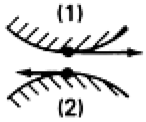
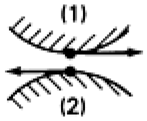
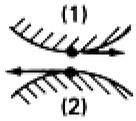
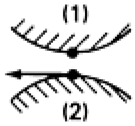
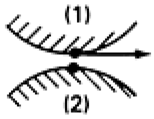
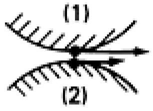
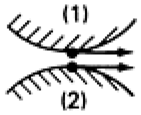
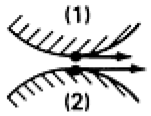
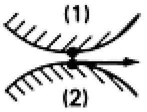
Operational Parameters

Operational parameters characterize the functional conditions of a tribosystem. They can be considered (with the exception of friction-induced temperatures) as independent variables that can be varied during tribological testing to obtain friction and wear data experimentally. The basic operational parameters in tribology are:

- *Type of motion,* that is, the kinematics of triboelements (1) and (2), to be classified in terms of sliding, rolling, spin, and impact and their possible superpositions (Fig. 3). The kinematics can be continuous, intermittent, reverse, or oscillating
- *Load* (F_N), defined as the total force (including weight) that acts perpendicular to the contact area between triboelement (1) and (2), as shown in Fig. 3
- *Velocity* (v), to be specified with respect to the vector components and the absolute values of the individual motions of triboelements (1) ad (2). According to the Table 2, distinctions must be made among the relative velocity v_r (relevant to friction-induced temperature rises), the sum velocity v_s (relevant, in lubricated tribosystems, to the formation of an elastohydrodynamic film), and the slide-to-roll ratios
- *Temperature* (T) of the structural components at stated location and time, that is, the initial (steady-state) temperature and the friction-induced temperature rise (average temperature rise and flash temperatures) to be estimated on the basis of friction heating calculations (see the following article in this Section on "Design of Friction and Wear Experiments")
- *Time* dependence of the set of operational parameters (F_N, v, T, for example, load cycles and heating or cooling intervals
- *Duration* (t) of operation, performance, or test

In addition to these functional operational parameters, disturbances such as external vibrations or radiation might need to be taken into consideration as well.

Table 2 Type of motion and velocities of the components of a tribosystem for sliding and sliding and rolling

Type of motion		Velocities u, v and slip s				
		u_1	u_2	$v_r = \|u_1 - u_2\|$	$v_s = \|u_1 + u_2\|$	$s = 2\left[\frac{u_1 - u_2}{u_1 + u_2}\right]$
Sliding						
Simple sliding		$u_1 > 0$	$u_2 = 0$	$v_r = u_1$	$v_s = u_1$	$s = 2$
Sliding		$u_1 > 0$ $\|u_1\| > \|u_2\|$	$u_2 < 0$	$v_r > u_1$	$v_s < u_1$	$2 < s < \infty$
Pure sliding		$u_1 > 0$	$u_2 = -u_1$	$v_r = 2u_1$	$v_s = 0$	$s = \pm\infty$
Sliding		$u_1 > 0$ $\|u_1\| < \|u_2\|$	$u_2 < 0$	$v_r > \|u_2\|$	$v_s < \|u_2\|$	$-\infty < s < -2$
Simple sliding		$u_1 = 0$	$u_2 < 0$	$v_r = \|u_2\|$	$v_s = \|u_2\|$	$s = -2$
Sliding and rolling						
Simple sliding		$u_1 > 0$	$u_2 = 0$	$v_r = u_1$	$v_s = u_1$	$s = 2$
Rolling with slip		$u_1 > 0$ $u_1 > u_2$	$u_2 > 0$	$v_r < u_1$	$v_s > u_1$	$0 < s < 2$
Pure rolling		$u_1 > 0$	$u_2 = u_1$	$v_r = 0$	$v_s = 2u_1$	$s = 0$
Rolling with slip		$u_1 > 0$ $u_1 < u_2$	$u_2 > 0$	$v_r < u_2$	$v_s > u_2$	$-2 < s < 0$
Simple sliding		$u_1 = 0$	$u_2 > 0$	$v_r = u_2$	$v_s = u_2$	$s = -2$

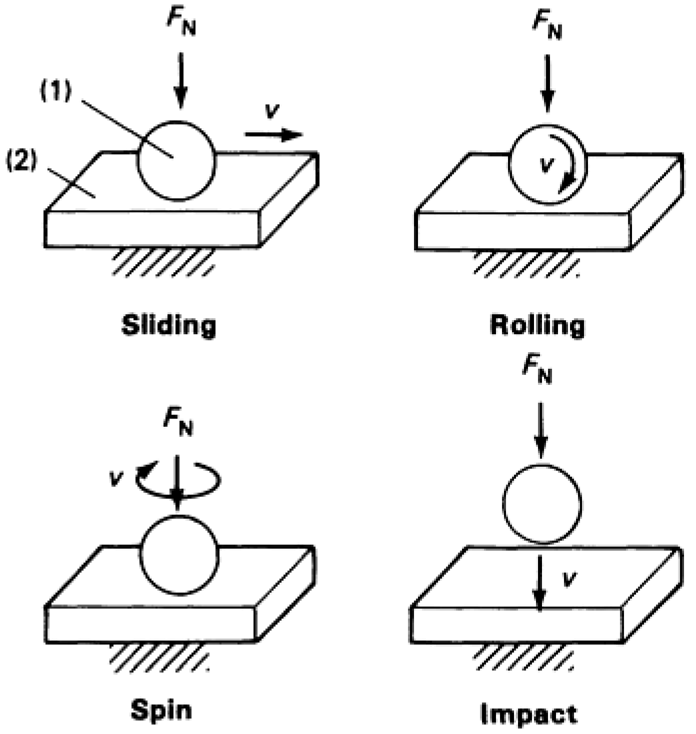


Fig. 3 Kinematics of tribosystems

Interaction Parameters

Interaction parameters characterize the action of the operational parameters on the structural components of tribosystems. These parameters define in particular the contact mode and the lubrication mode of a tribosystem with a given material/material or material/lubricant/material structure. The contact mode of two touching solid bodies is characterized microscopically by materials interactions, which are described by contact stresses and stress distributions. The materials and stress interactions cause a resistance against motion (friction) and may lead to surface damage (wear). Therefore, the materials and stress interactions in tribosystems are also called friction and wear mechanisms, or generally tribological processes, and specified in terms such as adhesion, abrasion, tribochemical reactions, surface fatigue, and so forth.

Interface Forces and Energies. Theoretically, the microscopic interaction forces between contacting solids include, at least in principle, all those types of atomic and molecular interaction that contribute to the cohesion of solids, such as metallic, covalent, and ionic, that is, primary chemical bonds (short-range forces), as well as secondary van der Waals bonds (long-range forces) (Ref 3, 4). These surface forces depend in a complicated manner on the physicochemical nature of the materials and the structure and composition of the outermost surface layers and contaminants. It should also be noted that the chemical composition, the electronic nature, and the microstructure of surfaces may be quite different from that of the subsurface (volume) of a material.

Experimentally, the only macroscopic way to characterize adhesive interactions between two solid bodies contacting under a normal load, F_N, is to destroy the bonding and to measure in the opposite direction to F_N the force, F_A, necessary for the separation of the surfaces. The ratio $a = F_A/F_N$ is termed the coefficient of adhesion. On the microscopic level, it is possible to determine with an atomic force microscope (noise level $\sim 2 \times 10^{-11}$ N) the interface forces (including friction forces) between single atoms of contacting surface tips (Ref 5).

In energetic terms, the formation of a solid/solid contact results in a net release of surface energy resulting from the replacement of two surfaces by one solid/solid interface of lower surface energy. The change in surface energy per unit area of contact, $\Delta\gamma$, can be written as:

$$\Delta\gamma = \gamma_1 + \gamma_2 - \gamma_{12} \quad \textbf{(Eq 1)}$$

where γ_1 is the surface energy of solid 1, γ_2 is the surface energy of solid 2, and γ_{12} is the surface energy of the interface.

Contact Deformation Modes. The surfaces of solid tribocomponents possess a certain roughness for which cross-sectional topography descriptors such as the centerline average (CLA), or the peak-to-valley roughness, as well as stochastic parameters such as the variances of the height and slope of asperities and the curvature are used for characterization (Ref 6). In addition, fractal geometry approaches may also be applied (Ref 7).

If two nominally flat and dry solid materials (1) and (2) are brought into static contact under the action of a normal load (Fig. 4), the touching asperities of this tribocontact deform elastically or plastically under the given load. The dominating contact deformation mode is governed by a deformation criterion (the so-called plasticity index), which depends on the deformation properties and the parameters of the surface topography of the contacting bodies (1) and (2) defined in Fig. 4 (Ref 6). The summation of individual contact spots gives the real area of contact, A_{or}, which usually is much smaller than the apparent geometrical area of contact. The real area of contact depends in the elastic case primarily on the ratio of F_N to the composite elastic modulus E' (see the discussion of "Contact Stresses" below) and in the plastic case on the ratio of F_N to the yield pressure, p_y, or the hardness, H, of the softer of the contacting bodies. If, in addition to the normal load, a friction force F_F is introduced, a junction growth of asperity contacts may occur, leading to a considerably larger area of contact. The basic parameters relevant to the contact mode of a materials/materials pair are summarized in Fig. 4. In engineering applications, the contact deformation mode may be additionally influenced by disturbances such as misalignments or vibrations.

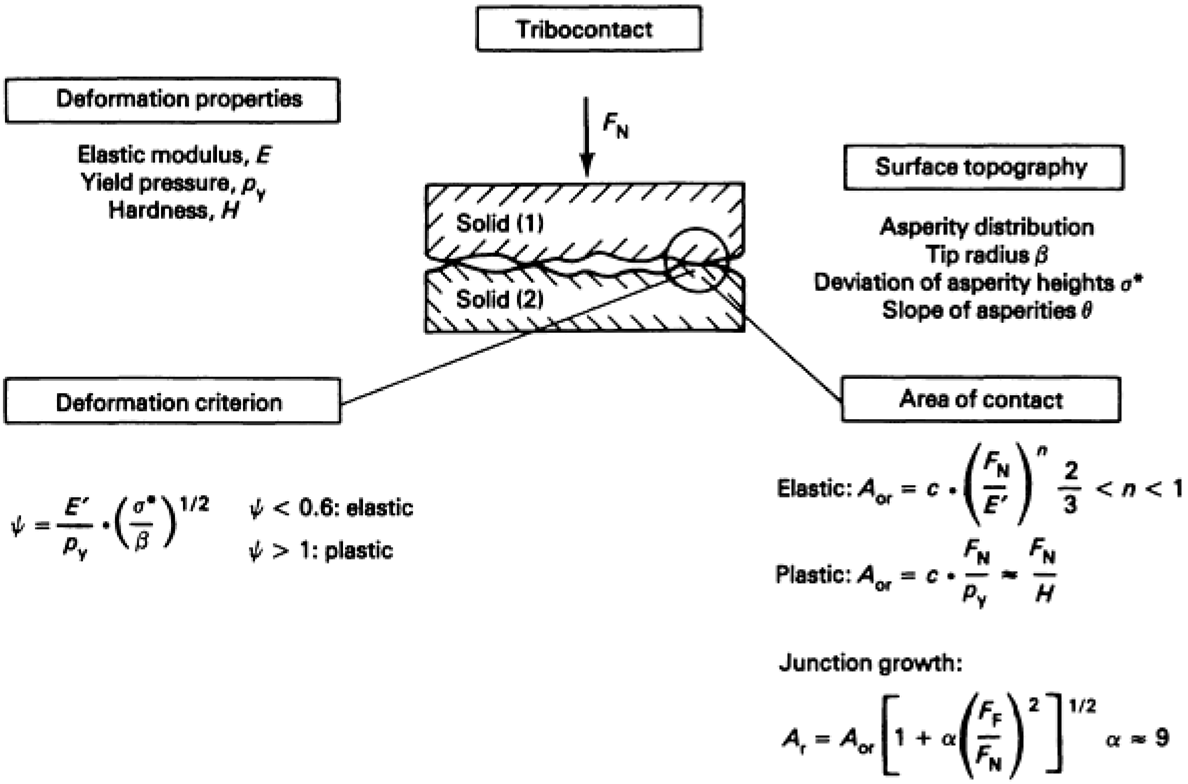


Fig. 4 Characteristics of a tribocontact between solid bodies

Contact stresses in a tribosystem depend on groups of parameters characterizing (1) the contact geometry, for example, conformal or contraformal contact configurations, (2) the elasticity, viscoelasticity, plasticity, or hardness of the materials, and (3) the external static or dynamic load.

For curved bodies, the macroscopic elastostatic contact situation is described by the Hertzian equations (Ref 8). To illustrate the basic parameters, consider the fundamental idealized case of the pure elastic contact of two spherical bodies (1) and (2) of radii r_1 and r_2, elastic moduli E_1 and E_2, and Poisson's ratios v_1 and v_2 under a normal load, F_N, as shown in Fig. 5(a). The contact pressure, p, at a location 1 within the contact area and the radius of contact a_H are given by:

$$p = \frac{3\,F_N}{2\pi a_H^2}\left[1 - \left(\frac{1}{a_H}\right)^2\right]^{1/2} \quad (0 \leq 1 \leq \alpha_H) \qquad \textbf{(Eq 2)}$$

$$a_H = \left[\frac{3r}{2E}\right]^{1/3} F_N{}^{1/3} \qquad \textbf{(Eq 3)}$$

where

$$r = \left[\frac{1}{r_1} + \frac{1}{r_2}\right]^{-1} \qquad \textbf{(Eq 4)}$$

is the equivalent radius of curvature, and

$$E = \left[\frac{1 - \nu_1^2}{2E_1} + \frac{1 - \nu_2^2}{2E_2}\right]^{-1} \qquad \textbf{(Eq 5)}$$

is the composite elastic modulus.

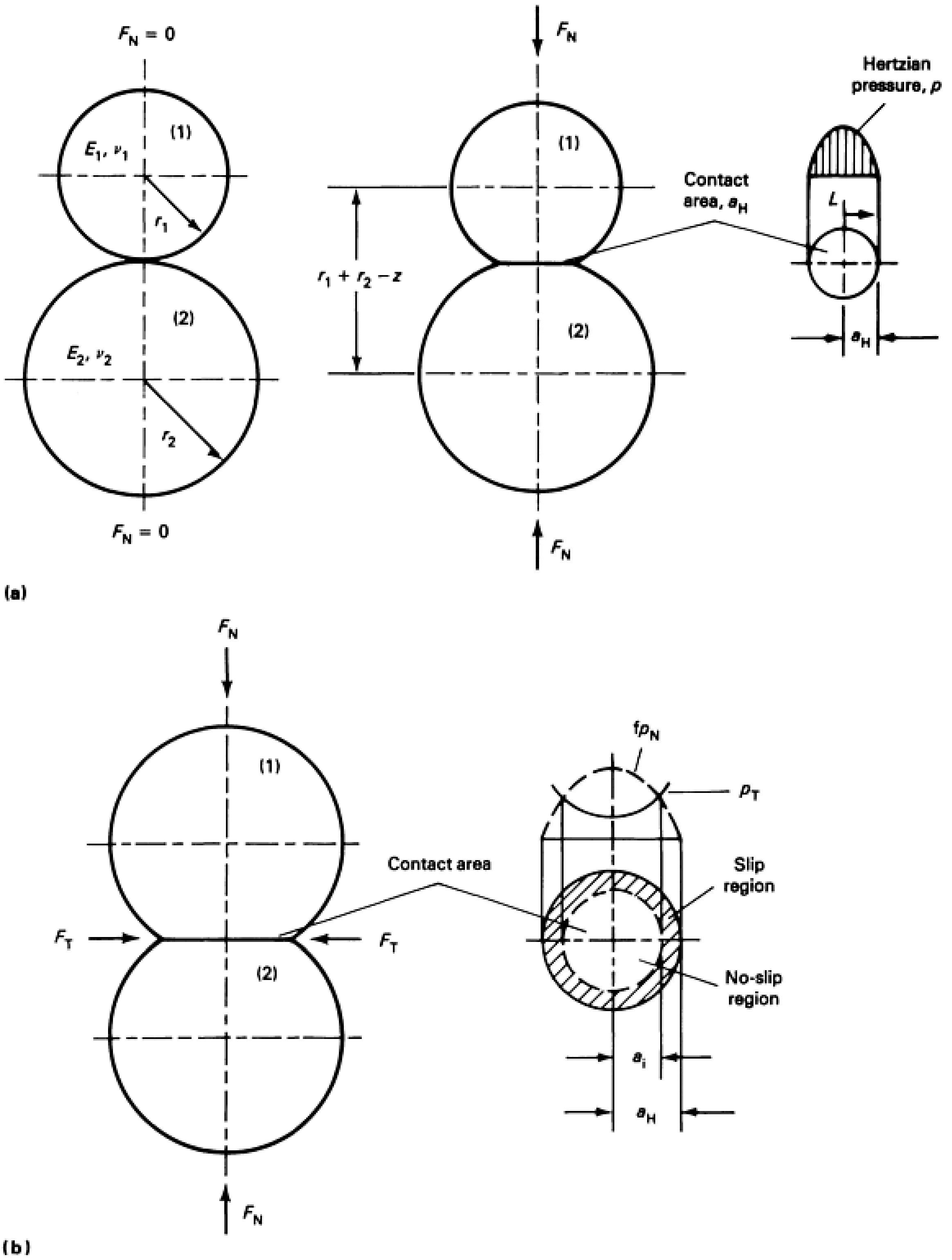


Fig. 5 Hertzian contact between spherical bodies under (a) normal forces and (b) combined normal and tangential forces

If a tangential force $F_T < F_N \cdot f_0$ (where f_0 is the static coefficient of friction) is additionally applied to a static Hertzian contact, slip and no-slip regions within the Hertzian contact area result before a macroscopic horizontal relative motion between body (1) and body (2) occurs. If a_H denotes the radius of the Hertzian contact circle, it follows, according to Mindlin (Ref 9), that no slip occurs within a circle of radius:

$$a_i = a_H \left[1 - \frac{F_N^2}{F_T^2}\right]^{1/3} \quad \textbf{(Eq 6)}$$

whereas in the part of the Hertzian contact zone in which the radius is between a_i and a_H, slip occurs (Fig. 5b).

In the case of macroscopic relative motion, the superposition of normal forces, F_N, and frictional forces $F_F > F_N \cdot f_0$ (perpendicular to F_N) in a tribocontact leads to complex stress distributions in the contacting bodies. The multiaxis stress condition may be converted into a uniaxial tensile stress in such a way that the uniaxial tensile load affects the material to the same extent as the multiaxial stress state. The basic stress hypotheses described below are suggested for tribocontacts such as rolling-element bearings or gears (Ref 10).

Distortion Energy Hypothesis (DEH). The equivalent stress, $\sigma_{E, DEH}$, is obtained by equating the distortion energies for uniaxial and multiaxial loading (Mises criterion):

$$\sigma_{E, DEH} = \frac{1}{2}^{1/2} [(\sigma_1 - \sigma_2)^2 + (\sigma_2 - \sigma_3)^2 + (\sigma_3 - \sigma_1)^2]^{1/2} \quad \textbf{(Eq 7)}$$

where $\sigma_1 > \sigma_2 > \sigma_3$ are the principal normal stresses.

The distortion energy hypothesis takes into account how all three differences between the principal stresses influence the initiation of plastic deformation of the material.

Shear Stress Hypothesis (SH). The extent to which the material is stressed is represented by the maximum effective principal shear stresses, T_{max} (Tresca criterion). Using the Mohr notation, the equivalent stresses are given by

$$\sigma_{E, SH} = |\sigma_1 - \sigma_3| = |2\, T_{max}| \quad \textbf{(Eq 8)}$$

where σ_1 and σ_2 are the principal normal stresses.

According to this hypothesis, fatigue occurs due to flow processes in the plane of the maximum principal shear stresses T_{max}, which is located at 45° to the principal stress system.

In addition to the contact pressure distribution of spherical bodies with a circular contact area illustrated in Fig. 5, Fig. 6 shows the contact pressure distribution in an elliptical Hertzian contact of two bodies with curvature in two planes. The resulting principal normal stresses σ_1, σ_2, σ_3 along the *z*-axis (directed into the material) for an axis ratio A/B = 6 are plotted on the left side. The equivalent stresses formed in accordance with the DEH and SH stress hypotheses are shown on the right.

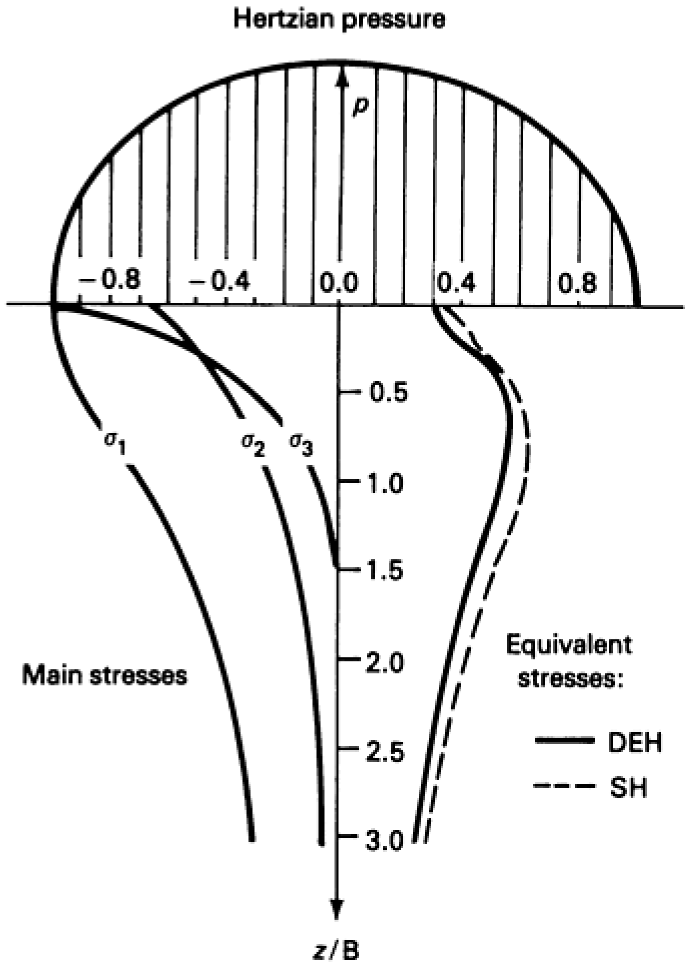
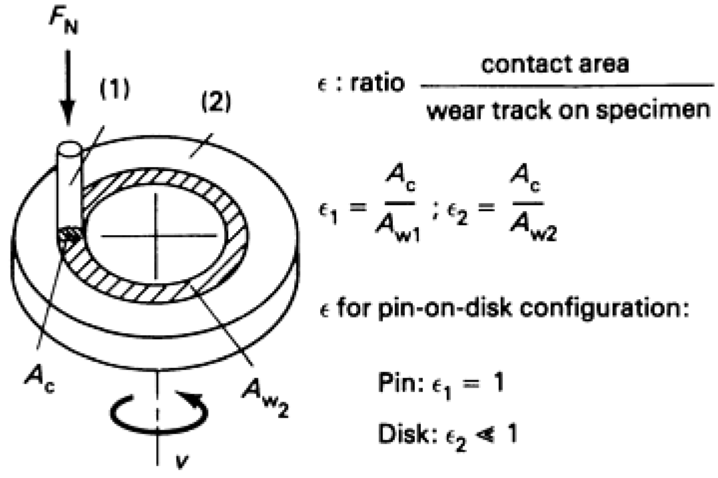


Fig. 6 Material stress for curved surfaces and elliptical contact (A/B = 6)

Contact Area/Wear-Track Ratio. The formation of a contact is a necessary condition for all friction and wear processes that occur, by definition, in the momentary contact area A_c. The kinematics and time-dependent variation of the contact area leads to wear tracks (A_{w1}, A_{w2}), which may be different for the tribocomponents (1) and (2) of a given tribosystem. As illustrated in Fig. 7 for the example of a pin-on-disk testing system, a tribocontact parameter, ϵ, can be defined. This tribocontact parameter is also called the contact area/wear-track ratio (Ref 11).



Characteristics of tribocomponents for which $\epsilon = 1$	Characteristics of tribocomponents for which $\epsilon < 1$
Permanent contact	Intermittent contact
No macroscopic cyclic stressing	Cyclic stressing
Permanent friction heating	Intermittent friction heating
Reduced material-atmosphere interactions	Direct material-atmosphere interactions on A_w - A_c

Fig. 7 The tribocontact parameter, ϵ, for a pin-on-disk configuration

If for a given tribosystem, the tribocontact parameters of tribocomponent (1) and tribocomponent (2) are unequal (that is, $\epsilon_1 \neq \epsilon_2$), then the two tribocomponents are subject to different tribological actions with respect to the kinematics, stresses, friction-induced heating, and material-atmosphere interactions as summarized in Fig. 7. It follows that in investigations of tribosystems, the response of both tribocomponents to the interfacial tribological processes must be carefully analyzed. Therefore, the tribocontact parameter plays an important role in the appropriate design of friction and wear test configurations to be used for simulative testing of the components of real triboengineering systems. For example, a material for a certain triboengineering component should only be tested in a test configuration where the materials specimens have the same tribocontact parameter as in the real triboengineering system.

The lubrication modes of tribosystems, consisting of triboelement (1), triboelement (2), and lubricant (3), are determined by structural parameters (for example, elastic moduli, E_1, and E_2, radii of curvature r_1 and r_2, surface roughness R_{a1} and R_{a2}, lubricant viscosity, μ, and the viscosity-pressure coefficient, α) and operational parameters (for example, load, F_N, velocity, v, and bulk oil temperature, T). Three main lubrication regimes can be distinguished by Stribeck curves (Ref 12). Figure 8 shows how lubrication regimes and variations of friction and wear coefficients are functions of the parameter combination $\mu \cdot v \cdot F_N^{-1}$, or the film thickness-to-roughness ratio λ.

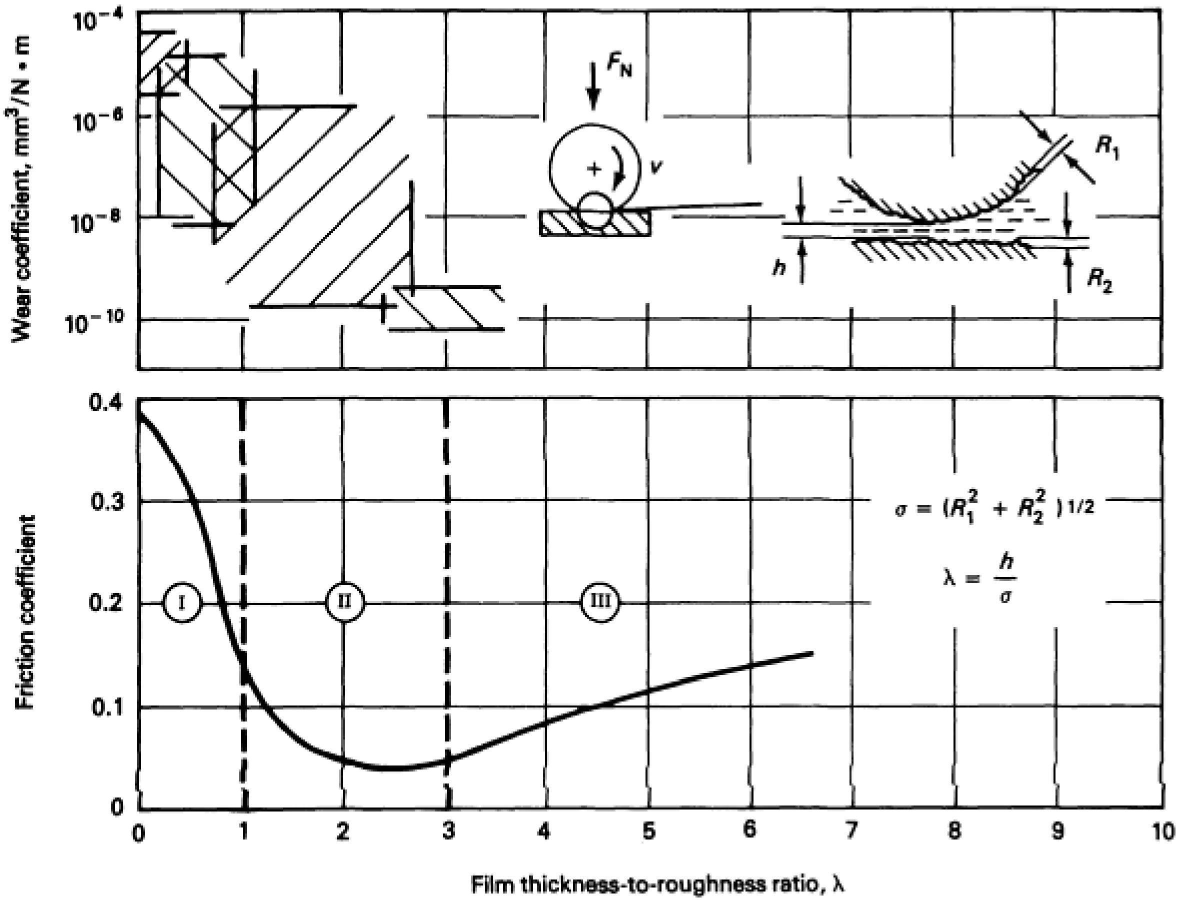


Fig. 8 Lubrication regimes and variations of friction and wear coefficients as functions of the ratio of film thickness to roughness

Regime I: boundary lubrication ($\lambda < 1$). In this regime, the tribological behavior is governed by solid/solid friction and wear processes, influenced by interface physics and chemistry.

Regime II: mixed lubrication or partial elastohydrodynamic (EHD) lubrication ($1 < \lambda < 3$). This regime consists of a dynamic, stochastically distributed coexistence between partial solid/solid asperity contacts and partial lubricant film separations.

Regime III: fluid film lubrication (hydrodynamic or EHD lubrication) ($\lambda > 3$). In this regime, the tribological behavior is determined by the rheology of the lubricant and can be calculated by the methods of fluid mechanics, for example, the Reynolds equation.

The different lubrication regimes are connected with different ranges of friction coefficient and wear coefficient values as illustrated schematically in Fig. 8. Additional information can be found in the article "Lubrication Regimes" in this Volume.

Tribometric Characteristics

Tribometric characteristics are measures of the results of interfacial interactions, that is, tribological processes in tribosystems. Friction and wear can be described phenomenologically by force-related or energy-related parameters with respect to friction, and by geometry-related or matter-related parameters with respect to wear (Ref 13). In addition to the numeric friction and wear parameters defined below, other tribometric characteristics, such as friction-induced noise or vibrations, may also be of interest. Further, the tribological processes as well as changes in the composition and microstructure of the triboelements and the resulting shape and composition of wear surfaces and wear particles must be characterized (see the Sections of this Handbook that deal with laboratory characterization techniques). It is important to

recognize, however, that friction and wear parameters are system-dependent characteristics that must be related to the pertinent tribosystem as described in the article "Presentation of Friction and Wear Data" in this Section.

Friction Parameters. Friction, usually classified as static or dynamic friction, is the resistance to initiate or sustain relative motion of contacting bodies (Fig. 1b). The diagnosis of the relevant parameters shows that depending on the kinematics (Fig. 3), friction can be measured as frictional force, F_F, in sliding and frictional torques, T_{FR} or T_{FSP}, in rolling or spinning, respectively (Ref 14). The static or dynamic sliding friction coefficient, f, is defined as the quotient between the friction force, F_F, to initiate or sustain relative motion tangential to the normal load, F_N, that is, $f = F_F/F_N$. It follows that a rolling friction coefficient and a spin friction coefficient can be defined as the quotient of the friction torque and the normal load times a characteristic radius related to the contact geometry. The frictional work or energy, E_F, with respect to the kinematics (Fig. 3) for sliding, rolling, and spinning, respectively, is given by:

$$E_{FS} = \int_{S_F} F_F \cdot ds_F \qquad \textbf{(Eq 9)}$$

where S_F is the sliding distance,

$$E_{FR} = \int_{\Psi_F} T_{FR} \cdot d\psi_F \qquad \textbf{(Eq 10)}$$

where ψ is the rolling angle, and

$$E_{FSP} = \int_{\Psi_{SP}} T_{FSP} \cdot d\psi_{SP} \qquad \textbf{(Eq 11)}$$

where ψ_{SP} is the spin angle.

In these equations the time dependence of the operational parameters has to be recognized (see the earlier section of this article on "Operational Parameters"). The average frictional power is defined as frictional work or energy divided by the operating duration, t (that is, $P_F = E_F/t$). For sliding friction, $P_F = F_F \cdot v = f \cdot F_N \cdot v$.

Wear Parameters. Wear is the progressive loss of substance from the operating surfaces of the mechanically interacting elements of tribo-systems. As illustrated in a simplified manner in Fig. 9, wear may be measured in terms of:

- *Length,* that is, one-dimensional changes in the geometry of interacting triboelements perpendicular to their common contact area
- *Area,* that is, two-dimensional changes of cross sections of interacting triboelements perpendicular to their common contact area
- *Volume,* that is, three-dimensional changes of geometric regions of interacting triboelements adjacent to their common contact area. Wear volumes are connected via density or specific gravity with wear masses or wear weights

In addition to these quantities, a wear-time ratio may be defined as wear velocity (Ref 15). Other common wear parameters are the wear rate, which is the wear volume per unit of sliding distance, and the wear coefficient, k, which is defined as:

$$k = \frac{W}{F_N \cdot s} \qquad \textbf{(Eq 12)}$$

where W is the wear volume (in mm^3), F_N is the applied load (in Newtons), and s is the sliding distance (in meters).

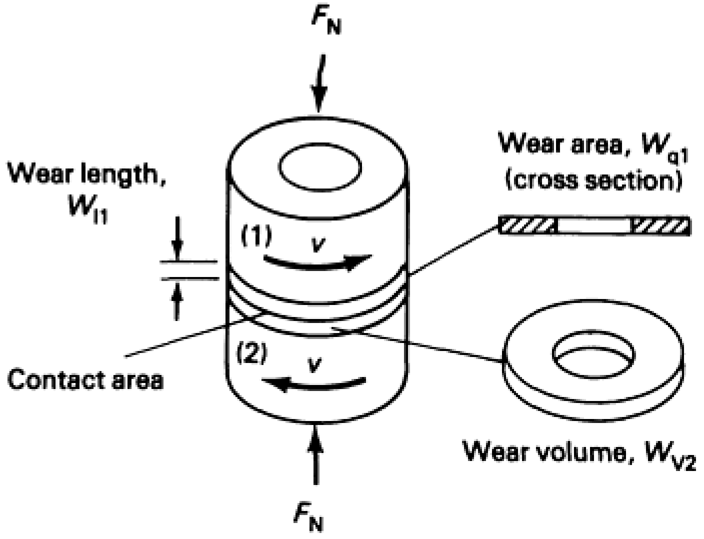


Fig. 9 Wear quantities illustrated for a tribosystem consisting of two sliding cylinders

It should be emphasized that wear quantities must be determined for both components involved in a wear process, that is, triboelement (1) and triboelement (2) of a given (closed) tribosystem.

Data Sheet of Basic Tribological Parameters

The foregoing analysis has shown that the diagnosis of the relevant parameters of tribosystems must recognize

- Structural parameters
- Operational parameters
- Interaction parameters
- Friction and wear parameters

A data sheet containing the main items of these parameter groups is given in Fig. 10. The data sheet, designed primarily for the characterization of closed tribosystems, can serve as a general guide for the compilation of the parameters to be specified individually for a given tribosystem. If necessary, the data sheet can be shortened or extended, but in any case at least the structural tribocomponents (1) to (4) must be identified and characterized and the set of the operational variables (F_N, v, T, t) determined.

Designation of tribo-system:
Function of tribo-system:

Structure / Properties	Tribo-element (1)	Tribo-element (2)	Lubricant (3)	Environment (4)
Designation				
Composition Density Volume				
Geometry Dimensions Roughness			Lubricants data (viscosity, etc):	Environments data (humidity, etc):
Modulus Hardness Toughness Microstructure				
Other				

Operational kinematics:			
Load F_N (N)		Velocity v(m/s)	
Contact area (mm^2)		Temperature T (°C)	
Contact pressure (N/mm^2)		Duration t (h)	

Contact area to wear track ratio	ε_1 (%):	ε_2 (%):
Film thickness to roughness ratio λ:		

Friction coefficient $\mu = F_F/F_N$	initial:	max:	final:

wear data / component	wear length W_1	wear area W_q	wear volume W_V	wear coefficient $k(mm^3/N \cdot m)$
tribo-element (1)				
tribo-element (2)				

Fig. 10 Data sheet for determining basic tribological parameters

In summary, the data sheet can support the following tasks:

- Systematic analysis of tribosystems and compilation of relevant parameters
- Guide for the planning and performance of tests or failure analyses
- Documentation of test results
- Input document for tribological databases

References

1. H. Czichos, *Tribology--A Systems Approach to the Science and Technology of Friction, Lubrication and Wear,* Elsevier, Amsterdam, 1978, p 351-353
2. "Wear: Terms, Systems Analysis of Wear Processes, Classification of the Field of Wear," DIN Standard 50 320, Beuth-Verlag, Berlin, Dec 1979
3. N.P. Suh, *Tribophysics,* Prentice-Hall, 1986, p 26-45
4. D.H. Buckley, *Surface Effects in Adhesion, Friction, Wear, and Lubrication,* Elsevier, Amsterdam, 1981, p 245-312
5. G.M. McClelland and S.R. Cohen, "Tribology at the Atomic Scale," Research Report RJ 7444, IBM, 26 Apr 1990
6. D. Tabor, A Simplified Account of Surface Topography and the Contact between Solids, *Wear,* Vol 32, 1975, p 269
7. A. Majumdar and B. Bushan, Role of Fractal Geometry in Roughness Characterization and Contact Mechanics of Surfaces, *J. Tribol.* (*Trans. ASME*), Paper 89-Trib-20
8. H. Hertz, On the Contact of Solid Elastic Bodies, *J. für die reine und angew. Mathem.,* Vol 92, 1881, p 156 (in German)
9. R.D. Mindlin, Compliance of Elastic Bodies in Contact, *J. Appl. Mech.* (*Trans. ASME*), Vol 71, 1949, p 259
10. E. Broszeit, T. Preussler, M. Wagner, and O. Zwirlein, Stress Hypotheses and Material Stresses in Hertzian Contact, *Z. Werkstofftech.,* Vol 17, 1986, p 238-246
11. H. Czichos, *Tribology--A Systems Approach to the Science and Technology of Friction, Lubrication and Wear,* Elsevier, Amsterdam, 1978, p 265-267
12. R. Stribeck, The Principal Properties of Sliding and Rolling Bearings, *VDI Zeitschrift,* Vol 46, 1902, p 1341, 1432, 1463 (in German)
13. P.J. Blau, The Units of Wear Revisited, *Lubr. Eng.,* Vol 45, 1989, p 609-614
14. "Friction in Bearings--Definitions, Types, Conditions, Physical Quantities," DIN Standard 50 281, Beuth-Verlag, Berlin, Oct 1977 (in German)
15. "Wear Measuring Quantities," DIN Standard 50 321, Beuth-Verlag, Berlin, Dec 1979 (in German)

Design of Friction and Wear Experiments

Horst Czichos, BAM (Germany)

Introduction

FRICTION AND WEAR EXPERIMENTS must be carefully designed in order to meet the objectives of the tribotesting to be performed. During the design process, the wide variety of (1) tribological phenomena in industrial applications, (2) mechanical engineering equipment, and (3) laboratory studies available for tribotesting must all be considered. Generally speaking, the scope of tribotesting can be classified into the following primary areas:

- Evaluation of the function, performance, maintainability, reliability, life, or efficiency of engineering tribosystems or tribocomponents
- Quality control of tribocomponents
- Characterization of the tribological behavior of materials and lubricants
- Investigation of basic tribological processes and friction-induced energy losses or wear-induced

materials losses

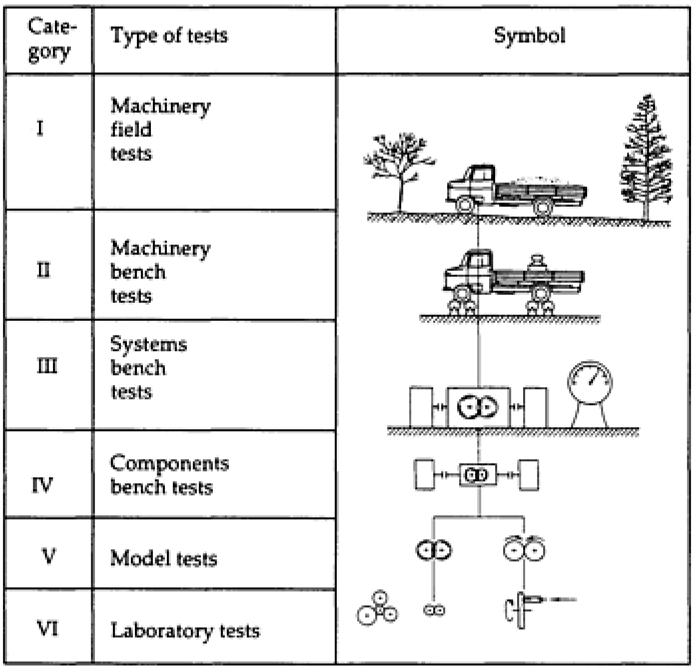
The tribotesting equipment to be used, the test conditions to be specified, and the friction and wear data to be measured must be selected with respect to the aims of tribotesting.

Categories and Conditions of Tribotests

Depending on the structure and function of the tribomachinery, tribosystem, tribocomponent, or specimen to be studied, tribotests can be grouped into six categories (Ref 1). These categories, which are listed in Fig. 1, consist of:

- *Category I: Machinery Field Tests.* Testing of actual tribomachinery under practical operating conditions
- *Category II: Machinery Bench Tests.* Testing of actual tribomachinery under practice-oriented (simplified, simulated, or accelerated) operating conditions
- *Category III: Systems Bench Tests.* Testing of specific tribosystems under practice-oriented operating conditions
- *Category IV: Components Bench Tests.* Testing of specific tribocomponents under practice-oriented operating conditions
- *Category V: Model Tests.* Testing of model test specimens under practice-oriented operating conditions
- *Category VI: Laboratory Tests.* Testing of arbitrary test specimens under laboratory operating conditions

Each tribotesting category has a different scope, and results obtained in one category cannot be simply transferred to another. This is obvious in considering the specific characteristics of the tribotesting categories listed in Fig. 1.



Category	Type of tests	Symbol
I	Machinery field tests	
II	Machinery bench tests	
III	Systems bench tests	
IV	Components bench tests	
V	Model tests	
VI	Laboratory tests	

Fig. 1 Categories of tribotesting

Tests of category I are performed with actual tribomachinery under practical conditions. Clearly, these tests represent the real performance of the object under study. However, full-scale field tests are usually very expensive. The broad

spectrum of the practical operating conditions often cannot be adequately characterized, and considerable efforts are necessary to confirm test results in a statistical manner in repeated tests.

In tests of category II to IV, in which actual tribomachinery, tribosystems, or tribocomponents are studied in well-defined bench tests, test results can be related to the actual triboengineering structures. An important consideration of these tests is the appropriate choice of operating conditions, which may be simplified, simulated, or accelerated as compared with the often unknown broad spectrum of the practical full-scale operating conditions. Sometimes attempts are made to determine the actual operating conditions of the full-scale filed performance with sensors and to simulate these conditions in less expensive bench tests. Because tribotests of categories I to IV differ in scope and testing conditions, the techniques and procedures to be applied for the determination of fraction and wear quantities must be related to the individual tests and cannot be generalized.

In tests of category V, model test specimens are used rather than actual tribocomponents. If an attempt is made to simulate actual triboengineering conditions with these tests, a sufficient similarity in the basic tribological parameters must be realized. These parameters, which were described in the previous article in this Section, include:

- The determination of the physicochemical nature of the model test specimens and the triboengineering system to be simulated (use of the same materials/lubricant/environment combination and connected materials properties)
- Similarity of the contact and lubrication mode (characterized, for example, by the tribocontact) parameter, ϵ, and the film thickness-to-roughness ratio, λ) to be obtained through an appropriate similarity in the kinematics, and the suitable choice of load, F_N, velocity, *v,* temperature, *T,* and test duration, *t,* as described in the article "Basic Tribological Parameters"

To support the appropriate selection of the structural and operational parameters of simulative testing, the characteristics of the actual tribosystem to be simulated should be compiled in a data sheet (see Fig. 10 in the article "Basic Tribologicial Parameters" in this Section), and the test conditions for simulative tests of category V should be selected accordingly.

Tests of category VI are used primarily in fundamental studies of friction and wear processes. The conditions of these tests are often selected to study specific tribological phenomena rather than to simulate real triboengineering behavior.

Laboratory Friction and Wear Tests and Simulative Tribotesting

The basic characteristics and relevant parameters of laboratory and simulative tests are shown in Fig. 2. The design of laboratory friction and wear tests should be carried out in the following steps:

- Choose a suitable test configuration for the test specimens of triboelement (1) and triboelement (2) and specify the geometry of the test configuration, materials characteristics and properties, and surface characteristics (clean surfaces before test)
- Characterize the interfacial element (3) (for example, the lubricant) and the environmental medium or atmosphere (4) in terms of their chemical nature, composition, and chemical and physical properties
- Choose a suitable set of the operational parameters, including type of motion, load (F_N), velocity (v), temperature (T), and test duration (t)
- Perform the tests as functions of varied structural parameters of the triboelements (for example, hardness or roughness) and operational parameters (for example, load cycles or velocity variations). The conditions of the tests, such as the time dependence of operational parameters, should be controlled by appropriate detectors or sensors and supported by on-line computer techniques
- Measure interesting tribometric characteristics, such as friction quantities, wear quantities, tribo-induced acoustic quantities (for example, noise or vibrations), and tribo-induced thermal quantities (for example, friction-induced temperature rise)
- Characterize the worn surfaces of triboelement (1) and triboelement (2) with respect to surface roughness and surface composition and structure

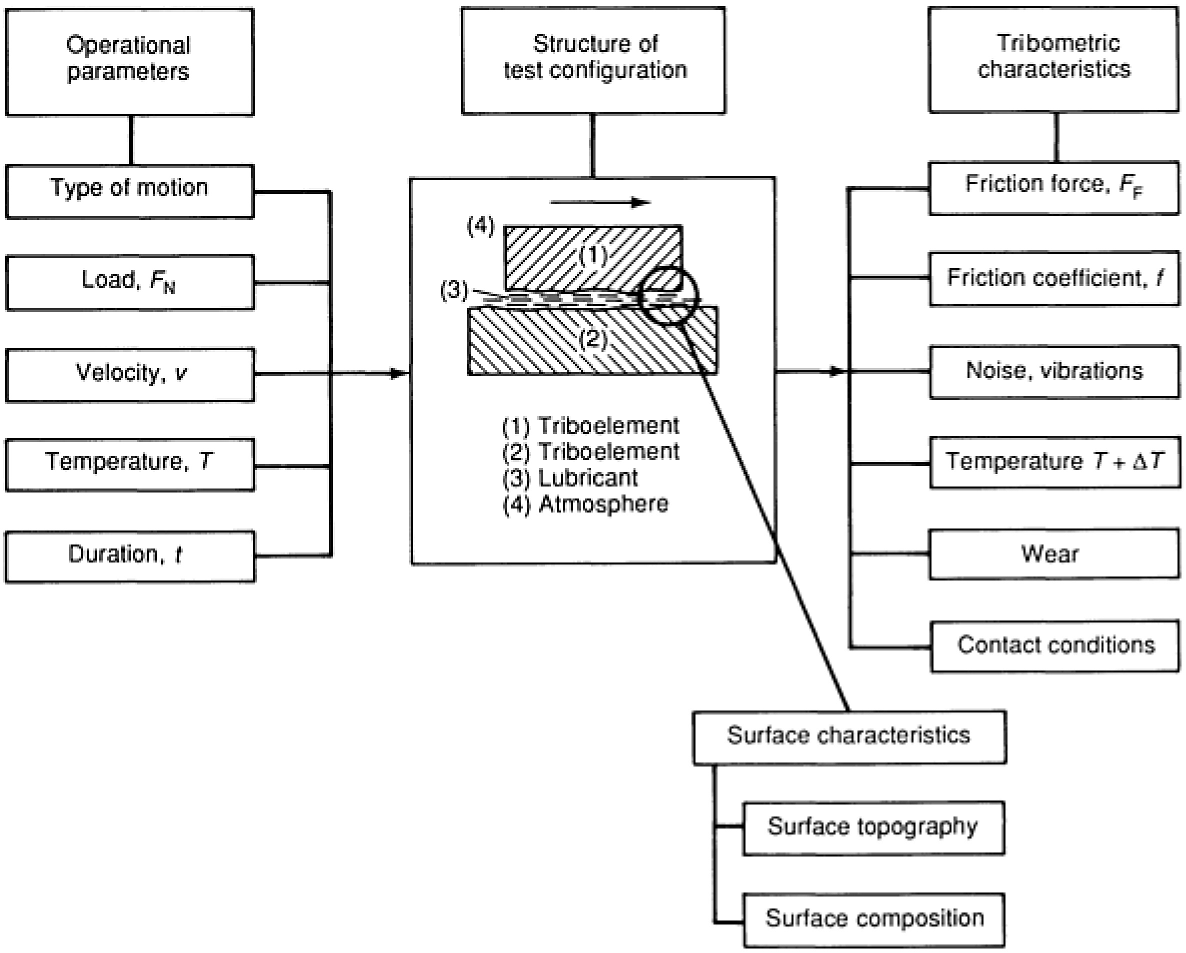


Fig. 2 Basic characteristics and parameters relevant to laboratory tribotesting

The results of laboratory tribotests, that is, the measured friction and wear quantities, can be symbolically represented by:

friction, wear = *f* (structural and operational parameters)

If an attempt is made to simulate the tribological behavior of a practical engineering tribosystem (TS) with a laboratory test configuration (TC), a sufficient similarity between the structural, operational, and interaction parameters of TC and TS must be reached. Figures 3(a), 3(b), 3(c), 3(d), and 3(e) show a series of tribotest flow charts used to support simulative tribotesting in the following steps:

- Selection of test configuration and structural parameters of test specimens (Fig. 3(a))
- Specification of operational parameters and specimen properties (Fig. 3(b))
- Tribotesting under unlubricated conditions (Fig. 3(c))
- Tribotesting under fluid-film lubrication (Fig. 3(d))
- Tribotesting under mixed or boundary lubrication (Fig. 3(e))

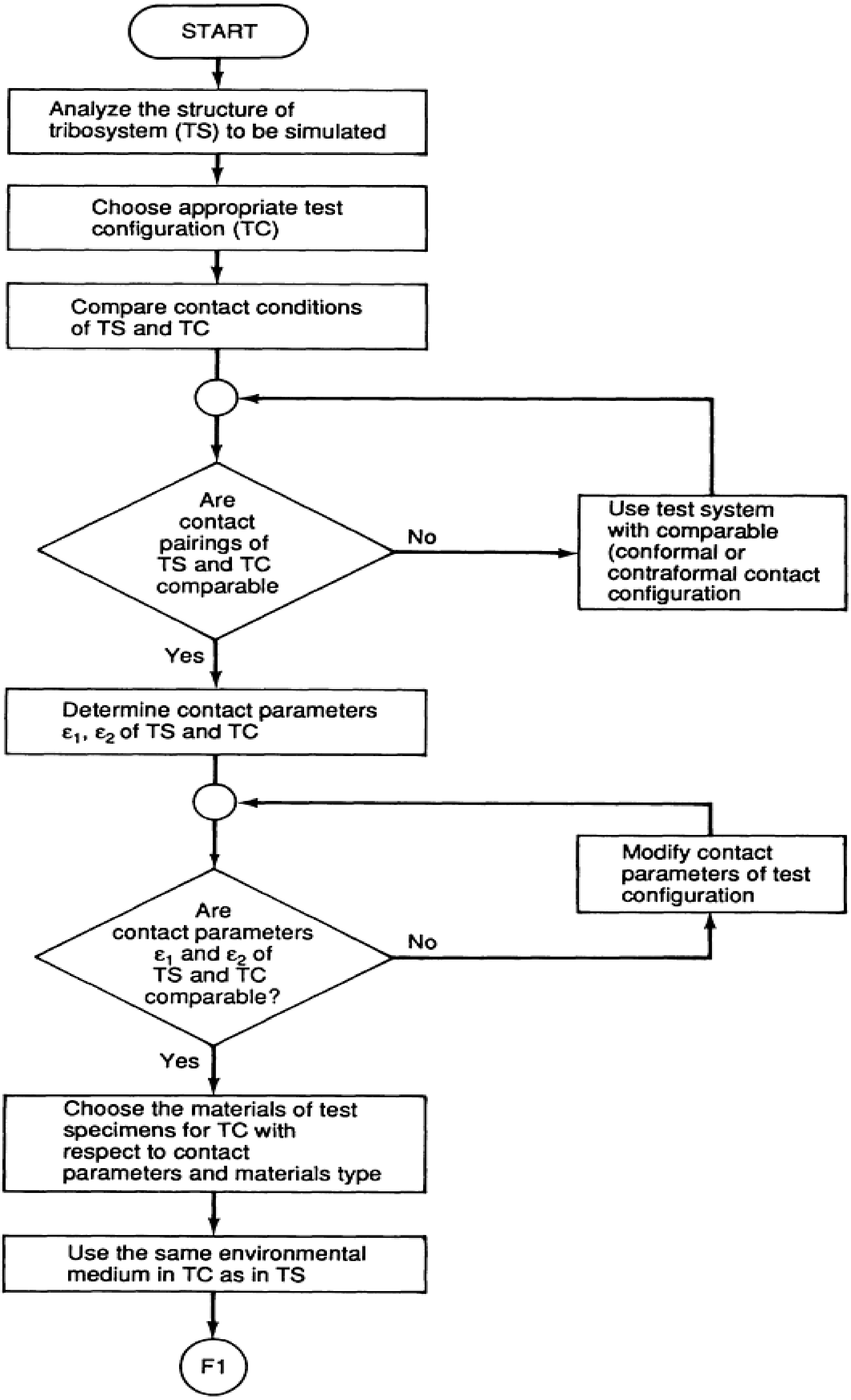


Fig. 3(a) Tribotest flow chart for selection of test configuration and structural parameters

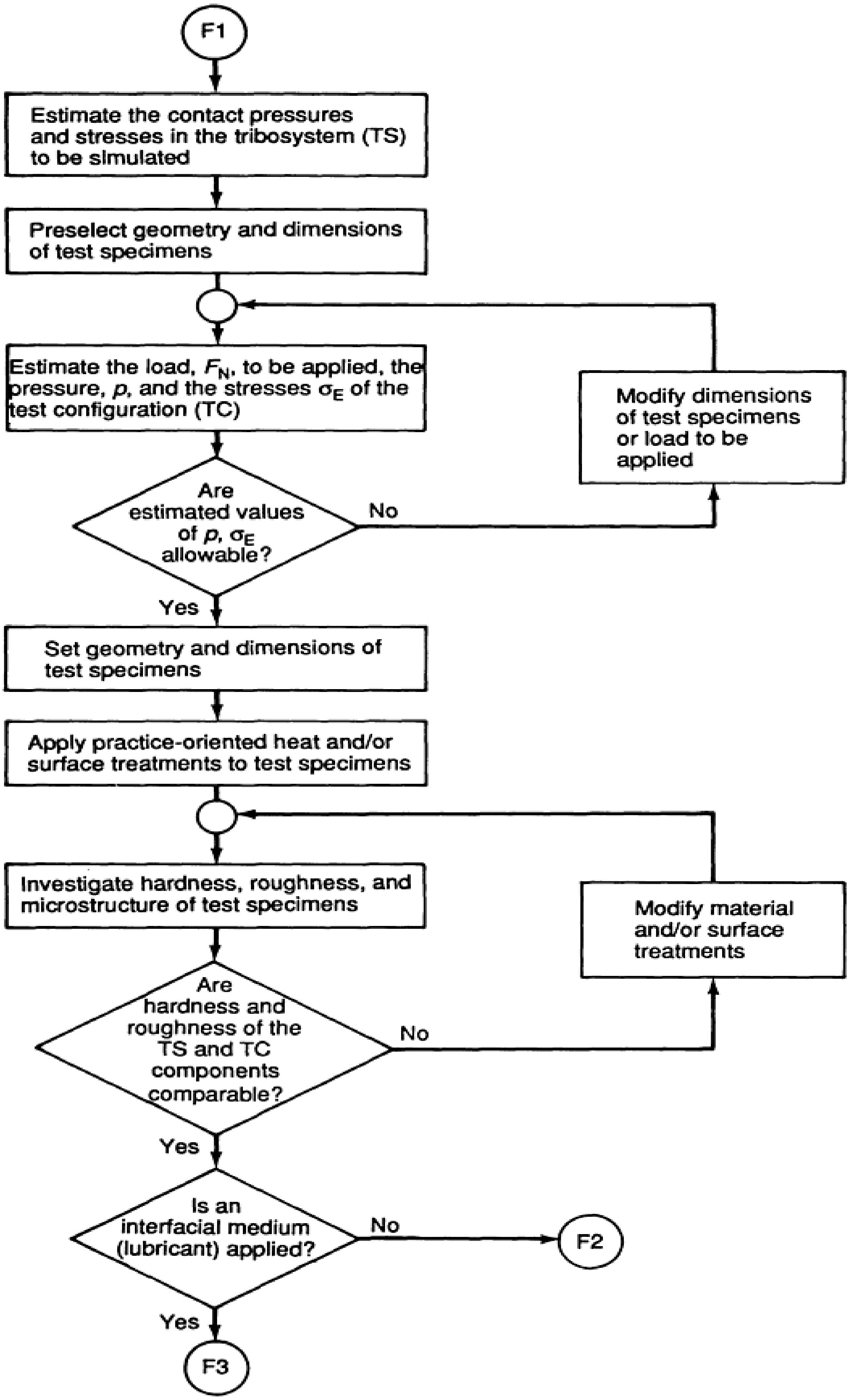


Fig. 3(b) Tribotest flow chart for specification of operational parameters and specimen properties

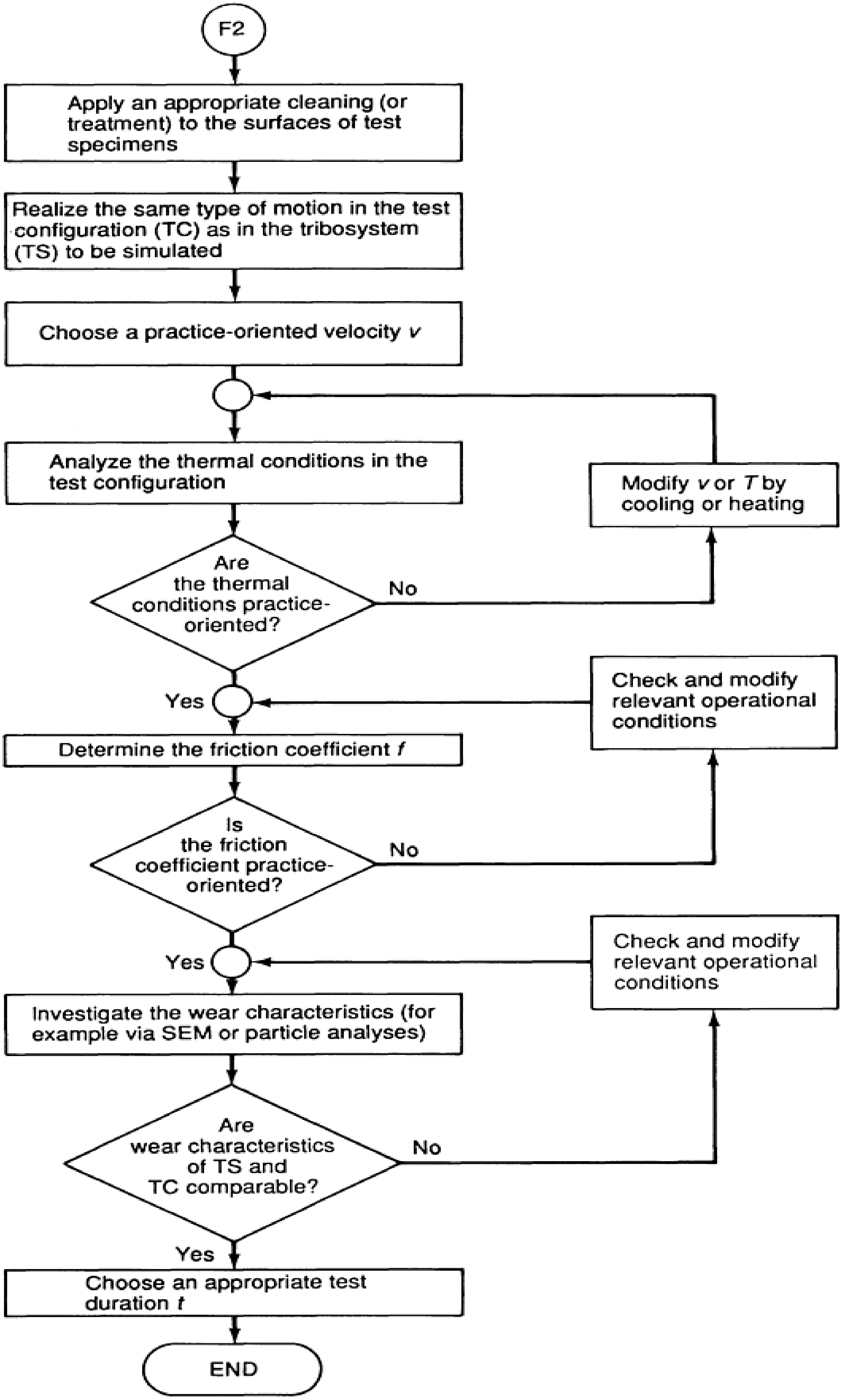


Fig. 3(c) Tribotest flow chart for testing under unlubricated conditions

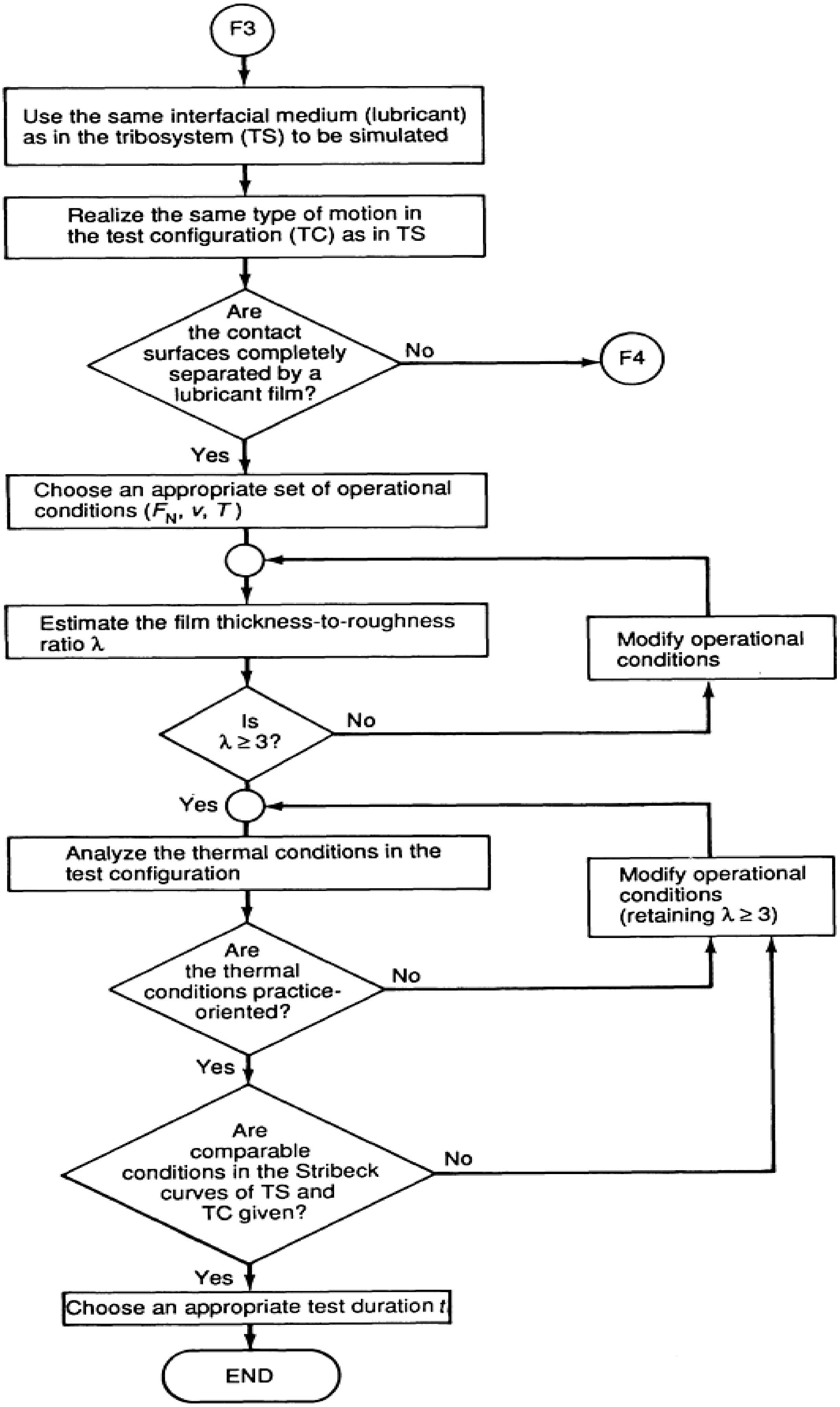


Fig. 3(d) Tribotest flow chart for testing under fluid-film lubrication

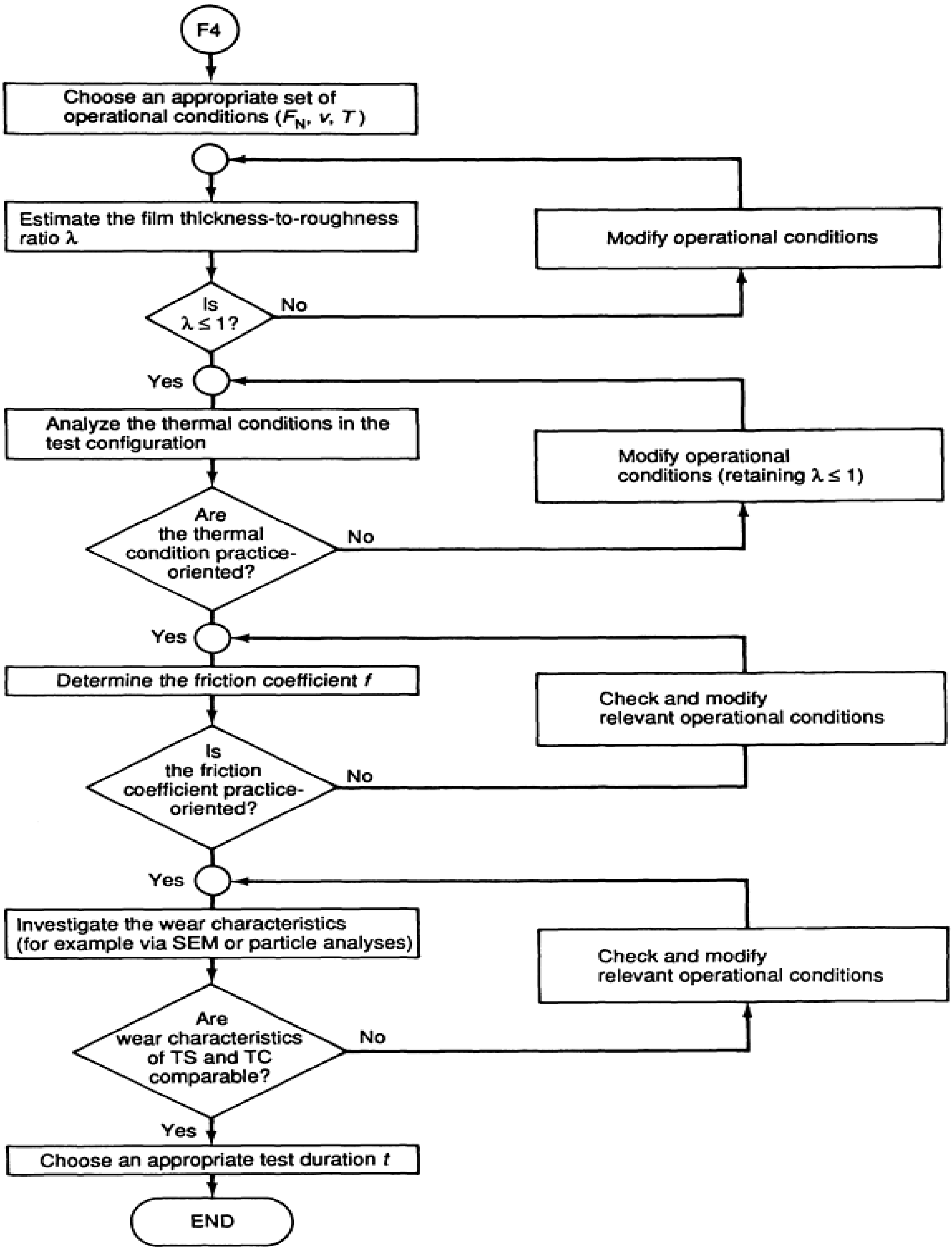


Fig. 3(e) Tribotest flow chart for testing under mixed or boundary lubrication

Evaluation of Tribotests

One of the primary problems in the evaluation of friction and wear data is coping with variability. Variability is inherent in every tribological process because of the great number of influencing parameters, their potential fluctuations, and time dependencies. Generally speaking, there are systematic deviations (bias) and random deviations to be considered in the evaluation of experimentally measured data. The two statistics most commonly used to characterize measured friction and wear data are the mean and the standard deviation (Ref 2). If y_1, y_2,. . ., y_n, are denoted the n individual observations of a random sample from a tribotest, the mean ($\bar{y}$) and standard deviation (s) are defined as follows:

$$\bar{y} = \sum_{i=1}^{n} y_i/n \qquad \textbf{(Eq 1)}$$

$$s = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n-1}} = \sqrt{\frac{\sum y_i^2 - \left(\sum y_i\right)^2/n}{n-1}} \qquad \textbf{(Eq 2)}$$

While the mean characterizes the location of the observed values on a corresponding scale, the standard deviation gives a measure of how the *n* observations vary or spread about the average. The square of the standard deviation is called variance.

Recall that $\bar{y}$ and *s* can be computed for a sample of *n* observations. Therefore, call $\bar{y}$ and *s* sample parameters the sample mean and sample standard deviation. It is important to note that these parameters can be understood as estimates of "ideal" parameters which are characteristics of an assumed population of observations or, in other words, a random variable, *n*, realizations of which form the sample of observations under consideration. Call the corresponding parameters population parameters, or more precisely, population mean and population standard deviation, and, denote them μ and σ, respectively. It is then well known that a sample parameter approximates the corresponding population parameter if *n* increases.

Let *y* be a sample mean and *s* a sample standard deviation obtained from *n* measurements of a tribological quantity. Assume that the distribution of the corresponding population is approximately normal. One can then specify a confidence interval for the population parameter μ, that is, an interval that contains μ with a given confidence level 1 - α, and a certain probability, *p* (for example, $\alpha = 0.05$ and $1 - \alpha = p = 0.95$) by:

$$\bar{y} - \frac{t \cdot s}{\sqrt{n}} \leq \mu \leq \bar{y} + \frac{t \cdot s}{\sqrt{n}} \qquad \textbf{(Eq 3)}$$

The values of *t* (depending on the number of measurements *n* and the confidence level 1 - α) can be taken from Table 1. It follows that the evaluation of tribotests may lead to the following results:

$$\text{Measurand} = \bar{y} \pm \frac{t \cdot s}{\sqrt{n}} \qquad \textbf{(Eq 4)}$$

where $\bar{y}$ and *s* are the sample parameters of a sample of *n* individual measurements of a friction or wear quantity.

Table 1 Chart of *t* values (*n* number of individual observations, 1 - α confidence level)

n	$\alpha = 0.10$ $t_{0.10}$	$\alpha = 0.05$ $t_{0.05}$	$\alpha = 0.01$ $t_{0.01}$
3	2.920	4.303	9.925
5	2.132	2.776	4.604
7	1.943	2.447	3.707
10	1.833	2.262	3.250
15	1.761	2.145	2.977

20	1.729	2.093	2.861
25	1.711	2.064	2.797
30	1.699	2.045	2.756
40	1.695	2.021	2.704
60	1.672	2.000	2.660
120	1.658	1.980	2.617
∞	1.645	1.960	2.576

In a similar manner, a confidence interval can be obtained for the population standard deviation σ:

$$s \cdot \sqrt{n-1} / \sqrt{\chi^2_{1-\alpha/2}} < \sigma < s \cdot \sqrt{n-1} / \sqrt{\chi^2_{\alpha/2}} \quad \textbf{(Eq 5)}$$

where the values of χ^2 (depending on *n* - 1 and α) can be taken from the χ^2 distribution (Table 2).

Table 2 Chart of x^2 values (*n* number of individual observations, 1 - αconfidence level)

n	α= 0.10		α= 0.05		α= 0.01	
	$\chi^2_{1-\alpha/2}$	$\chi^2_{\alpha/2}$	$\chi^2_{1-\alpha/2}$	$\chi^2_{\alpha/2}$	$\chi^2_{1-\alpha/2}$	$\chi^2_{\alpha/2}$
3	5.991	0.103	7.378	0.051	10.597	0.010
5	9.448	0.711	11.143	0.484	14.860	0.207
7	12.592	1.635	14.449	1.237	18.548	0.676
10	16.919	3.325	19.023	2.700	23.589	1.735
15	23.685	6.571	26.119	5.629	31.319	4.075
20	30.144	10.117	32.852	8.907	38.582	6.844
25	36.415	13.848	39.364	12.401	45.559	9.886
30	42.557	17.708	45.722	16.047	52.336	13.121
40	54.572	25.695	58.120	23.654	65.476	19.996
60	77.931	42.339	82.117	39.662	90.715	34.770

Design of Experiments

Mathematical models, computer simulations, and statistical techniques are of fundamental importance for the appropriate design of experiments. Because detailed information on the tools used to design experiments is beyond the scope of this article, only a minimum of approaches will be outlined. For a more extensive treatment of the design and analysis of experiments, the reader is referred to Ref 3 and 4.

One-at-a-Time Approach. In the conventional approach to model or laboratory tribotesting, one quantity (*y*), for example, a friction quantity or a wear quantity, is measured as a function of only one independent variable (x_1) to be selected out of the following three categories: (1) test duration, *t*, (2) an operational variable (for example, load), or (3) a structural variable (for example, material hardness), where all other parameters and conditions are kept constant. In varying the values of the independent variable (x_1), a relationship of the form $y = f(x_1)$ is obtained. In a second test, the dependence of the measured quantity (*y*) on another independent variable (x_2) may be determined. The results of both tests can be jointly represented as a three-dimensional $y = f(x_1,x_2)$ graph.

An advantage of such one-at-a-time experiments is that the influence of single independent variables on the measured quantity can be clearly understood. However, a large number of experiments are necessary to determine the simultaneous influence of more than one variable on the measured quantity. For example, 24 experiments were needed to show the influence of the basic set of operating variables (pressure *p*, velocity *v*, temperature *T*) on the friction coefficient of a dry sliding polytetrafluoroethylene (PTFE)/materials pair, as shown in Fig. 4 (Ref 5).

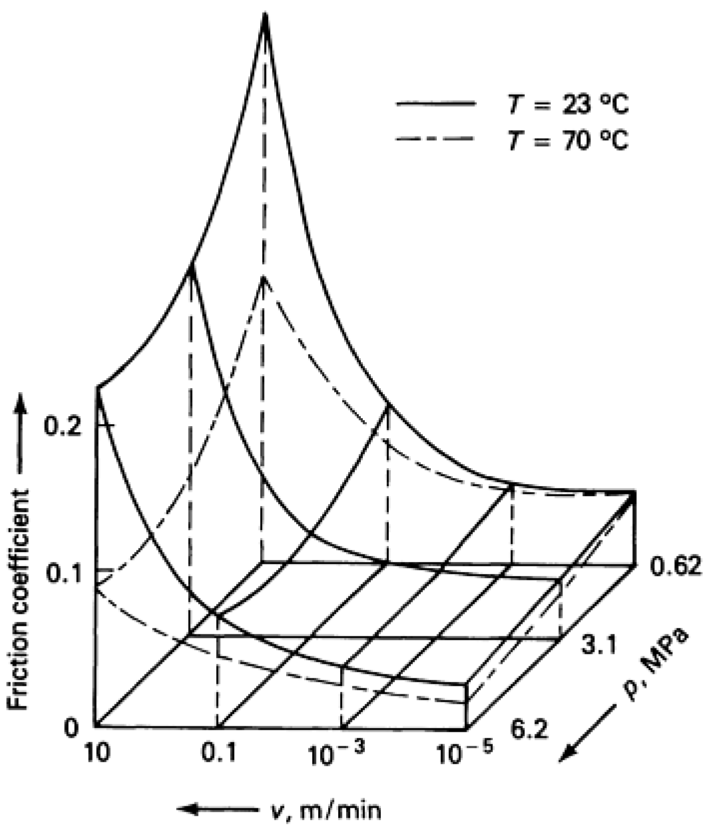


Fig. 4 Friction coefficient of PTFE/steel (AISI 52100, 800 HV, R_a = 0.02 μm) as a function of pressure (*p*), sliding velocity, (*v*), and temperature (*T*)

Simple Complete Factorials Approach. A complete factorial is a design of experiments in which a measured quantity is determined under the full set of combinations of the levels of all factors, that is, the variables included (Ref 4). It allows an assessment of the main effects of the single factors and the interaction effects of all factors. Although this method is based on the assumption of a linear dependence between the measured quantity and each of the independent variables, it may also serve as an approximative technique for nonlinear conditions as in tribology (Ref 6). The approach of the complete factorial design of experiments will be explained below by using data obtained in the example of tribotesting shown in Fig. 4.

The aim is to assess the main effects and the interaction effects of the three operational variables (pressure *p*, velocity *v*, and temperature *T*), each at two levels (denoted as high, h and low 1), on the measured friction coefficient *f*. The design is called a 2^3 factorial, and the eight operating variable combinations can be displayed graphically as a cube (Fig. 5). The values of the friction coefficient measured in the eight tests are listed in Table 3. The average value of the friction coefficient is:

$$\bar{f} = (0.05 + 0.03 + 0.36 + 0.22 + 0.05 + 0.02 + 0.18 + 0.09)/8 = 0.125 \quad \textbf{(Eq 6)}$$

Table 3 Test conditions and measured friction data

No. of test	Level of (*pvT*)	Pressure (*p*), MPa	Velocity (*v*), m/min	Temperature (*T*), °C	Friction coefficient (*f*)
1	(lll)	0.62	10^{-5}	23	0.05
2	(hll)	6.2	10^{-5}	23	0.03
3	(lhl)	0.62	10	23	0.36
4	(hhl)	6.2	10	23	0.22
5	(llh)	0.62	10^{-5}	70	0.05

6	(hlh)	6.2	10^{-5}	70	0.02
7	(lhh)	0.62	10	70	0.18
8	(hhh)	6.2	10	70	0.09

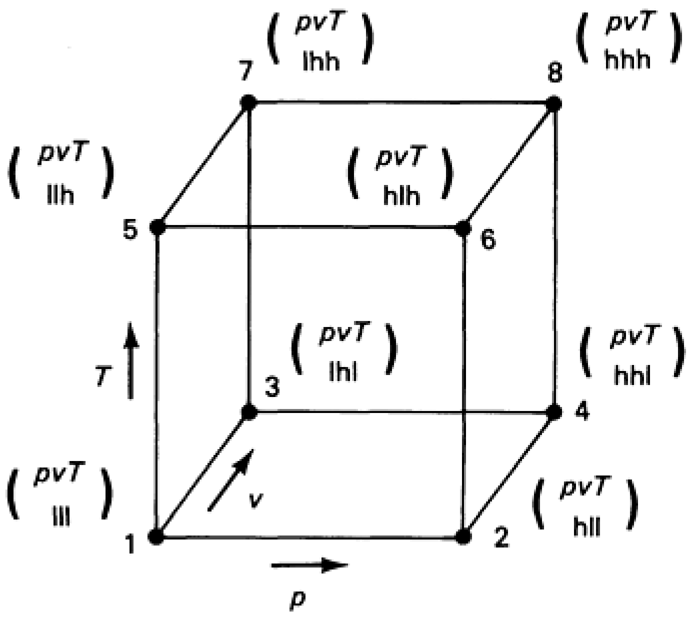


Fig. 5 Geometrical representation of the operating conditions of a 2^3 factorial design

By using the eight treatment combinations, the corresponding measured values of *f,* and the main effects of pressure, velocity, and temperature *T,* the two-factor interactions *pv, pT, vT* and the three-factor interaction *pvT* can be estimated using the data in Table 3.

Effect of Pressure. The effect of *p* when *v* and *T* are at low levels is (2 - 1) as shown in Fig. 5. Similarly, the effect of *p* when *v* is at the high level and *T* is at the low level is (4 - 3).

The effect of *p* when *T* is at the high level and *v* is at the low level is (6 - 5). Finally, the effect of *p* when both *v* and *T* are at the high level is (8 - 7).

Thus, the average effect of *p* is the average of these four combinations, or:

$$\Delta f(p) = (2 - 1 + 4 - 3 + 6 - 5 + 8 - 7)/4 \quad \textbf{(Eq 7)}$$

Equation 7 can be developed as a contrast between the four treatment combinations in the right face of the cube of Fig. 5 (where *p* is at the high level) and the four in the left face (where *p* is at the low level). Combining these gives:

$$\Delta f(p) = (2 + 4 + 6 + 8 - 1 - 3 - 5 - 7)/4$$
$$\Delta f(p) = -0.070 \quad \textbf{(Eq 8)}$$

which means that an increase of the contact pressure in this example lowers the average friction coefficient.

The effect of velocity v is a contrast between the four treatment combinations in the front face of the cube in Fig. 5 and the four in the back. This yields:

$$\Delta f(v) = (3 + 4 + 7 + 8 - 1 - 2 - 5 - 6)/4$$
$$\Delta f(v) = +0.175 \quad \textbf{(Eq 9)}$$

which means that an increase of the velocity increases the average friction coefficient considerably.

The effect of temperature T is a contrast between the four treatment combinations in the top face of the cube in Fig. 5 and the four in the bottom, that is:

$$\Delta f(T) = (5 + 6 + 7 + 8 - 1 - 2 - 3 - 4)/4$$
$$\Delta f(T) = -0.08 \qquad \textbf{(Eq 10)}$$

which means that an increase in the temperature lowers the average friction coefficient (similar to an increase of *p*).

Interaction Effects of Pressure, Velocity, and Temperature. When *T* is at the low level, the effect of the *pv* interaction is the average difference in the *p* effect at the two levels of *v*, that is, [(4 - 3) - (2 - 1)]/2. When *T* is at the high level, the *pv* interaction is [(8 - 7) - (6 - 5)]/2. The average *pv* effect is the average of these two combinations, or:

$$\Delta f(p,v) = (4 - 3 - 2 + 1 + 8 - 7 - 6 + 5)/4$$
$$\Delta f(p,v) = -0.045 \qquad \textbf{(Eq 11)}$$

which shows that the additional influence of *v* weakens the effect of *p* as discussed above in the description of effect of pressure.

The average effects of *pT* and *vT* are:

$$\Delta f(p,T) = (1 - 2 + 3 - 4 - 5 + 6 - 7 + 8)/4$$
$$\Delta f(p,T) = +0.01$$
$$\Delta f(v,T) = (1 + 2 - 3 - 4 - 5 - 6 + 7 + 8)/4$$
$$\Delta f(v,T) = -0.075 \qquad \textbf{(Eq 12)}$$

Equation 12 indicates that the combined effects of *p* and *T* in this example are almost negligible, whereas the effect of *T* at the two levels of *v* leads to a considerable decrease of the friction coefficient.

The *pvT* effect is defined as the average difference between the *pv* interaction for the two different levels of *T*. Thus:

$$\Delta f(p,v,T) = (8 - 7 - 6 + 5 - 4 + 3 + 2 - 1)/4$$
$$\Delta f(p,v,T) = +0.015 \qquad \textbf{(Eq 13)}$$

This combined effect can be neglected.

Summarizing the steps illustrated in the above discussion, a 2^3 factorial design, with 8 experiments to be performed, can be used to determine the effect of three independent variables (x_1, x_2, x_3), each at two levels, on a measured quantity *y* in tribotesting. Table 4 summarizes the results of the example discussed above, allowing an assessment of the main effects $\Delta f(p)$, $\Delta f(v)$, $\Delta f(T)$, and of the interaction effects $\Delta f(p,v)$, $\Delta f(p,T)$, $\Delta f(v,T)$, $\Delta f(p,v,T)$ on the coefficient of friction. Extended tables for more than three independent variables, or variation of these variables at more than two levels, can be designed accordingly (Ref 7, 8).

Table 4 Scheme for calculating effects in the 2^3 design of experiments

No. of test	Level of (x_1x_2,x_3)	$\bar{y}$	$\Delta y(x_1)$	$\Delta y(x_2)$	$\Delta y(x_1,x_2)$	$\Delta y(x_3)$	$\Delta y(x_1,x_3)$	$\Delta y(x_2,x_3)$	$\Delta y(x_1,x_2,x_3)$
1	(lll)	+1	-1	-1	+1	-1	+1	+1	-1
2	(hll)	+2	+2	-2	-2	-2	-2	+2	+2
3	(lhl)	+3	-3	+3	-3	-3	+3	-8	+3
4	(hhl)	+4	+4	+4	+4	-4	-4	-4	-4
5	(llh)	+5	-5	-5	+5	+5	-5	-5	+5
6	(hlh)	+6	+6	-6	-6	+6	+6	-6	-6
7	(lhh)	+7	-7	+7	-7	+7	-7	+7	-7
8	(lhh)	+8	+8	+8	+8	+8	+8	+8	+8
Result:		$\sum$/8	$\sum$/4	$\sum$/4	$\sum$/4	$\sum$/4	$\sum$/4	$\sum$/4	$\sum$/4

Example (see Table 3)	$\bar{f}$	$\Delta f(p)$	$\Delta f(v)$	$\Delta f(p,v)$	$\Delta f(T)$	$\Delta f(p,T)$	$\Delta f(v,T)$	$\Delta f(p,v,T)$
	0.125	-0.07	+0.175	-0.045	-0.08	+0.01	-0.075	+0.015

Round-Robin Tests

Terminology. The following qualitative terms (Ref 9) are of key relevance for the planning and evaluation of round-robin tests. More detailed information on the quantitative aspects of round-robin tests can be found in Ref 10, 11, 12.

Precision is a general term for the closeness of agreement between randomly selected individual measurements or test results. The standard deviation of the error of measurement may be used as a measure of "imprecision."

Repeatability is the closeness of agreement between mutually independent replicate test results obtained with the same test method on identical test materials, under the same conditions (that is, the same operator, laboratory, apparatus, and time).

Reproducibility is the closeness of agreement between replicate test results obtained with the same test method on identical test materials but under different conditions (that is, different operator, laboratory, apparatus, or time).

General Rules for Round-Robin Tests. A standard practice for conducting an interlaboratory study to determine the precision of a test method is given in ASTM E 691 (Ref 11). The general rules for the planning of round-robin tests are described in Ref 10. According to these standards, the aspects listed below are of key importance for the performance of round-robin tests.

Scope of the Tests. The goals to be reached during the round-robin tests should be determined from the outset.

Planning and Organization of the Tests. All participating partners should be informed of the scope of the test. Personnel for the test should include a coordinator of the entire test, a statistician for the planning and evaluation, and operators for the performance of the tests.

Guiding the Performance of the Tests. A clear description of the test method to be applied, the characterization of test materials, the test conditions, the variables to be varied, the quantities to be measured, and the number of repeated tests to be performed must be given. In addition, the format of compiling and describing the obtained results is to be specified.

Analyzing the Test Results. The specimens should be carefully selected, prepared, and characterized for the tests. The quantities to be measured and the format of reporting the measured data must be specified. For the analysis and evaluation of the test results as part of the standard round-robin test procedure, well-defined statistical techniques must be applied.

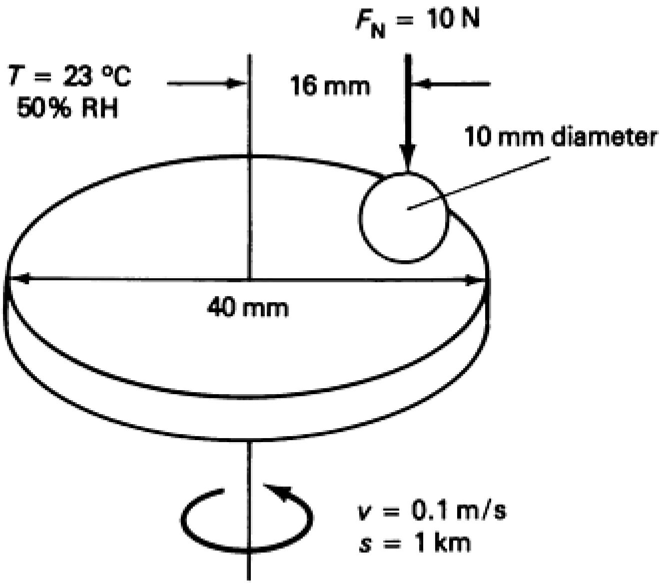
Case Study: International Round-Robin Sliding Wear Tests. An international round-robin comparison on the comparability, repeatability, and reproducibility of friction and wear tests has been performed within the framework of the Versailles Project on Advanced Materials and Standards (VAMAS) (Ref 13). Under the guidance of an international

committee on wear test methods, round-robin studies with α-Al_2O_3 ceramic and steel specimens were performed with 31 participating institutions from seven countries.

For the sake of simplicity, a ball-on-disk specimen configuration was selected as the test system. The test specimens were uniformly manufactured with respect to their geometry, dimensions, and surface finish. The conditions outlined below were specified on the basis of a thorough systems analysis of the main relevant characteristics of the tribotesters.

Test System. Conditions for the test system, which featured the ball-on-disk configuration shown in Fig. 6, were as follows:

- Stationary ball (10 mm, or 0.4 in., in diameter) against rotating disk (40 mm, or 1.6 in., in outer diameter and 32 mm, or 1.3 in., in track diameter)
- Rotation in the horizontal plane
- Direction of rotation of disk to be indicated by each laboratory
- Ball, disk, and debris to be collected and protectively stored in plastic containers
- Holders for disk and ball are left to the discretion of each laboratory
- Participants were asked to report (if possible) vibrations (for example, vibration amplitudes and frequency distribution) of the test rig at the stated location
- Participants were asked to report (if available) the stiffness data of the test rig



Material pairings			Measurements
Disk \ Ball	Steel	Alumina	• Friction force
Steel	Kit 1	Kit 2	• Wear (system, ball, disk)
Alumina	Kit 3	Kit 4	• Wear surfaces (SEM, profilometry)

Fig. 6 Conditions of VAMAS round-robin sliding wear tests. F_N, normal load; *v*, sliding velocity; *s*, sliding distance; *T*, ambient temperature; RH, relative humidity of the atmosphere

Materials used during the tests included an α-Al_2O_3 ceramic and AISI 52100 steel. The composition, roughness, hardness, and microstructure data were provided to each test participant.

Atmosphere. Laboratory air with a relative humidity (RH) of 50 ± 10% and a temperature of 23 ± 1 °C (73 ± 2 °F) was to be used.

Lubricant. No lubricant was to be used in the round-robin tests.

Operating variables included:

- Motion, continuous unidirectional sliding
- Normal load, F_N = 10 N (2.25 lbf)
- Sliding velocity, v = 0.1 m/s (0.33 ft/s)
- Temperature, T = 23 ± 1 °C (73 ± 2 °F)
- Sliding distance, s = 1 km (0.62 mile)
- Number of tests, 3 to 5

Preparation of Surfaces. Specimens were to be used in the as-received condition, that is, no mechanical surface finishing by the participants was necessary because specimens were centrally prepared in a uniform manner prior to distribution

Surfaces were cleaned immediately prior to each test. Washing in freon ($Cl_2FC\text{-}CF_2Cl$) was preferred to washing in ethyl alcohol since the water content in the alcohol could corrode the metallic surfaces. Following washing, the specimens were dried in warm air, rinsed with hexane, and re-dried in a drying oven at 110 °C (230 °F) for 30 min.

In addition, only chemicals of pure quality were to be used and samples had to be stored and transported in desiccators.

Wear and Friction Measurements. For wear measurements, each participant was asked to indicate whether wear of ball, wear of disk, or total wear of both ball and disk were measured. Participants were also required to continuously measure and record linear wear, to weigh specimens before and after each test, to measure the wear scar diameter on the ball with an optical microscope, and to provide profilogram results of both surfaces before and after the test.

Similarly for friction measurements, each participant was asked to indicate whether the frictional force or the friction torque were measured. They were also required to submit a simplified graph giving the fluctuations at the beginning and the end of the test, as well as minimum and maximum deviations during the test.

For both wear and friction measurements, all debris was collected.

Examination. Surfaces and debris were to be examined by optical color photography and scanning electron microscopy techniques. All photomicrographs had to be supplied at the standard magnifications of 50×, 100×, 200×, 500×, and 1000×.

A detailed instruction sheet containing the above points and specimen pairings ready to be tested were sent from a central source to all participants.

Data Evaluation. For the evaluation of numeric friction and wear data, the results of the participating laboratories were treated in the following manner to obtain statistically sound data.

Only the results of laboratories that performed more than one test run per specimen kit were considered. If laboratories reported only average values (as estimated by themselves from repeated test runs) these values were treated as if obtained from three measurements.

From the results obtained from the participating laboratories, the following four types of numeric data were determined:

- The friction coefficient was determined after a sliding distance of 1000 m (3300 ft)
- The wear rate of the system was measured under steady-state conditions, that is, the displacement of the test specimens perpendicular to the sliding interface between 300 and 1000 m (1000 and 3300 ft) sliding distance divided by the sliding distance. It should be noted that this is a geometry-dependent quantity

that cannot be compared directly with the wear rates of other test systems

- The ball wear scar diameter
- The disk wear-track width. From the numeric data of the four quantities (friction coefficient, wear rate, ball wear scar diameter, and disk wear-track width), the mean values and the standard deviations for the different material pairings were calculated (Table 5)

Table 5 Results of the VAMAS round-robin sliding wear tests.

See Fig. 6 for test conditions.

	Kit 1 Steel-steel	Kit 2 Ceramic-steel	Kit 3 Steel-ceramic	Kit 4 Ceramic-ceramic
Coefficient of friction[(a)]	0.60 ± 0.11	0.76 ± 0.14	0.60 ± 0.12	0.41 ± 0.08
Number of data	109	75	64	76
Number of laboratories	26	26	23	26
Wear rate of system, μm · km^{-1}[(b)]	70 ± 20	Very small	81 ± 29	Very small
Number of data	47	. . .	29	. . .
Number of laboratories	11	. . .	11	. . .
Ball wear scar diameter, mm	2.11 ± 0.27	[(c)]	2.08 ± 0.35	0.3 ± 0.05
Number of data	102	. . .	60	56
Number of laboratories	23	. . .	21	19
Disk wear-track width, mm	[(d)]	0.64 ± 0.13	[(d)]	Not measured
Number of data	. . .	54	. . .	. . .
Number of laboratories	. . .	19	. . .	. . .

(a) At 1000 m sliding distance.

(b) Determined from the wear curve (steady-state range between 300 and 1000 m sliding distance).

(c) Material transfer from disk to ball.

(d) Material transfer from ball to disk.

A comparison of the reported data showed that all the laboratories were able to perform the tests under the specified operating conditions of load, velocity, and ambient temperature. However, the values of the relative humidity of the ambient atmosphere varied between 13% and 78% and less than half of the participating laboratories were able to keep the operating conditions within the specified range of 50 ± 10% RH.

The results of the VAMAS round-robin comparison show that good reproducibility for the numeric friction and wear data was obtained. These data can be summarized in terms of the relative standard deviations (s_r and s_R according to ASTM Standard E 691 divided by the mean value) as follows (Ref 13):

- Reproducibility of friction data: within laboratories ±9% to ±13%; interlaboratory, ±18% to ±20%
- Reproducibility of specimen wear data: within laboratories, ±5% to ±7%; interlaboratory, ±15% to ±20%
- Reproducibility of systems wear data: within laboratories, ±14%; interlaboratory, ±29% to ±38%

These results indicate that the reproducibility of friction and wear measurements is comparable with that of other engineering quantities, provided that the tests are performed under well-controlled conditions, applying standard round-robin test procedures.

Appendix: Calculations of Elastic Contact Dimensions and Stresses (Ref 14)

Peter J. Blau, Oak Ridge National Laboratory

General parameters used to calculate the elastic contact dimension and the stress of nonconforming surfaces are given in Table 6. The relation of these variables to specific contact geometries (Fig. 7) is given in Table 7.

Table 6 Elastic contact and stress nomenclature

Symbol	Definition
P	Normal force (load)
p	Normal force (load) per unit contact length
E_1	Modulus of elasticity (Young's modulus) for body 1
E_2	Modulus of elasticity (Young's modulus) for body 2
v_1	Poisson's ratio for body 1
v_2	Poisson's ratio for body 2
D	Diameter when only one body in the contact is curved
D_1	Diameter of body 1 (in the case of two curved bodies, by convention $D_1 > D_2$)
S_c	Maximum compressive stress (Hertzian stress)
a	Radius of the elastic, circular contact
b	Width of the contact (used for cylinders)
c	Semimajor axis of an elliptical contact area (used for crossed-cylinders)
d	Semiminor axis of an elliptical contact area (used for crossed-cylinders)
α	Geometric factor used in crossed-cylinders calculations
β	Geometric factor used in crossed-cylinders calculations

Table 7 Formulas for calculating elastic contact dimensions and stresses of contact geometries shown in Fig. 7

Geometry of the contact	Elastic contact dimension[(a)]	Maximum compressive stress
Sphere-on-flat	$a = 0.721\ ^3\sqrt{PDC_0}$	$S_c = 0.918\ ^3\sqrt{P/(D^2C_0^2)}$
Cylinder-on-flat	$b = 1.6\ \sqrt{pDC_0}$	$S_c = 0.798\ \sqrt{p/(DC_0)}$
Cylinder-on-cylinder (parallel axes)	$b = 1.6\ \sqrt{pC_0/\mathrm{A}}$	$S_c = 0.798\ \sqrt{(pA)/C_0}$
Ball-in-spherical socket	$a = 0.721^3\ \sqrt{(PC_0)/B}$	$S_c = 0.918\ ^3\sqrt{P(B/C_0)^2}$
Cylinder-in-a-circular groove	$b = 1.6\ \sqrt{(pC_0)/B}$	$S_c = 0.798\ \sqrt{(pB)/\mathrm{C}_0)}$
Crossed-cylinders[(b)] (axes perpendicular)	$c = \alpha^3\sqrt{(PC_0)/A}$ $d = \beta_c$	$S_c = 0.4775\ (P/cd)$

(a) $C_0 = [(1 - v_1^2)/E_1] + [(1 - v_2^2)/E_2)$; $A = [(D_1 + D_2)/D_1D_2]$; $B = [(D_1 - D_2)/D_1D_2]$.

(b) To determine the values of α and β, first divide the larger diameter by the smaller to obtain a diameter ratio, then use Fig. 8 to determine the appropriate values for the two constants.

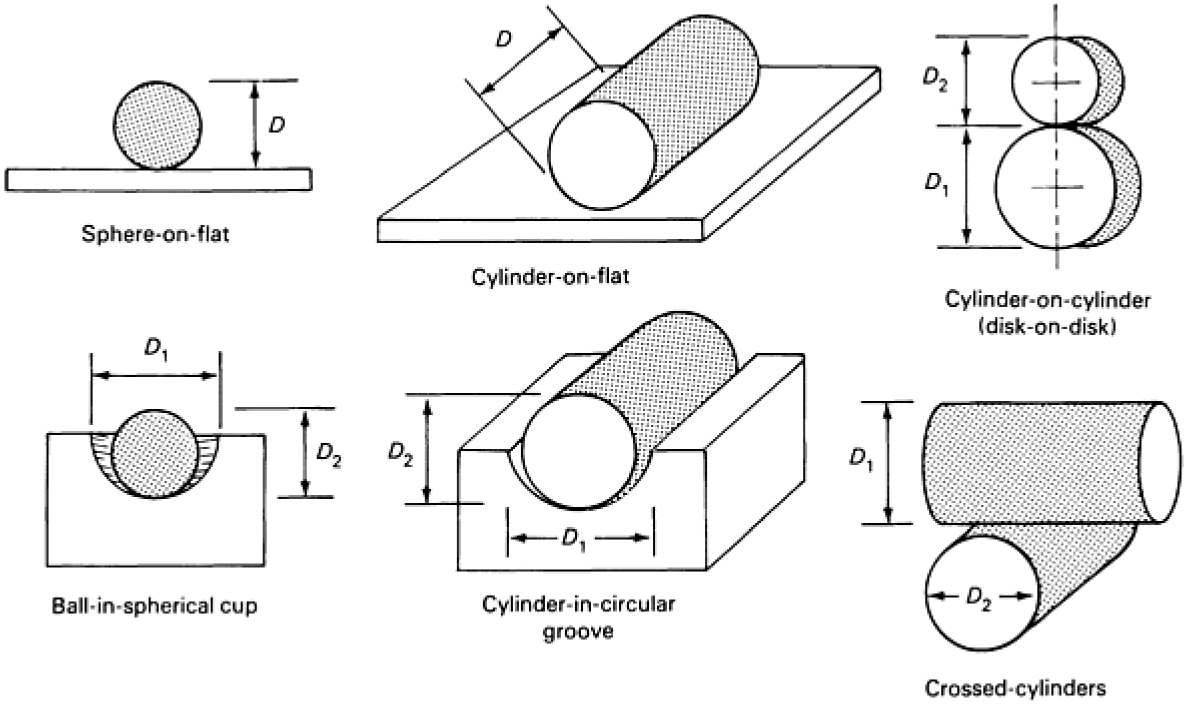


Fig. 7 Typical contact geometry models showing relation of *D, D_1*, and *D_2* in elastic contact and stress analysis (see Table 7)

To calculate the maximum stress level of crossed-cylinder geometry designs, an additional step is required. Two elastic contact dimension constants, α and β, must be obtained from Fig. 8 to calculate *c* and *d,* which are then substituted into the required formula for maximum compressive stress.

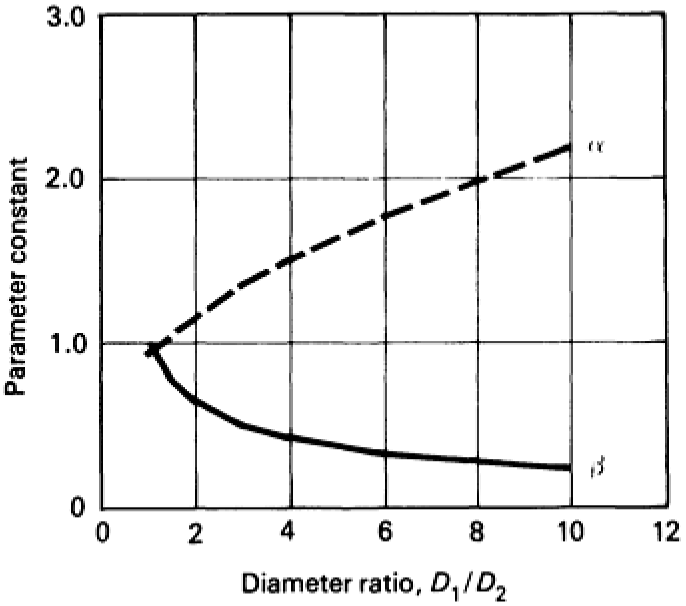


Fig. 8 Value of parameter constants, α and β, as a function of the ratio of diameters (*D_1/D_2*, where *D_1* > *D_2*) for crossed-cylinder contact geometries

References

1. "Wear--Specification of the Categories of Wear Testing," DIN Standard 50 322, Beuth-Verlag, Aug 1984 (in German)
2. R.W. Kennard, Engineering Statistics and Quality Control, *Marks' Standard Handbook for Mechanical Engineers,* McGraw-Hill, 1978, p 17-19
3. O.L. Davies, *Design and Analysis of Industrial Experiments,* Oliver and Boyd, Edinburgh, 1971, p 261-263
4. P.W.M. John, *Statistical Design and Analysis of Experiments,* Macmillan, London, 1971
5. H. Czichos, *Tribology--A Systems Approach to the Science and Technology of Friction, Lubrication and Wear,* Elsevier, Amsterdam, 1978, p 316-321
6. W. Mücke, Application of Statistical Planning of Experiments in Tribology, *Schmierungstechnik,* Vol 11, 1980, p 140-143 (in German)
7. D.C. Montgomery, *Design and Analysis of Experiments,* John Wiley, 1978, p 180-214
8. S. Kotz, N.L. Johnson, and C.B. Read, *Encyclopedia of Statistical Sciences,* John Wiley, 1982, p 359-366
9. "Standard Terminology Relating to Statistics," E 456, *Annual Book of ASTM Standards,* ASTM
10. "Precision of Test Methods--Determination of Repeatability and Reproducibility for a Standard Test Method by Inter-Laboratory Tests," Standard 5725, International Organization for Standardization, Geneva, 1986
11. "Standard Practice for Conducting an Inter-laboratory Study to Determine the Precision of a Test Method," E 691, *Annual Book of ASTM Standards,* ASTM
12. W. Gerisch and D. Amthor, A Computer Program for the Planning of Round Robin Precision Experiments in the Case of the Random Effects Two-Way Layout, Based on Multipliers of Mandel and Lashof, *VDI Forschungsh.,* Vol 662, VDI-Verlag, Düsseldorf, 1990
13. H. Czichos, S. Becker, and J. Lexow, Multilaboratory Tribotesting: Results from the Versailles Advanced Materials and Standards Programme on Wear Test Methods, *Wear,* Vol 114, 1987, p 109-130
14. W.C. Young, *Roark's Formulas for Stress and Strain,* 6th ed., McGraw-Hill Book Company, 1989

Presentation of Friction and Wear Data

Horst Czichos, BAM (Germany)

Introduction

THE RESULTS OF TRIBOTESTS lead to friction and wear data, generally referred to as tribodata, which are system-dependent characteristics presented in the form of tribographs, transition diagrams, or tribomaps.

Tribographs are graphical presentations of a measured friction or wear quantity as function of:

- Operational parameters, for example, load (F_N), velocity (v), temperature (T), or test duration (t)
- Structural parameters, for example, materials pairings, hardness, roughness, or microstructure
- Interaction parameters, for example, contact stresses, film thickness-to-roughness ratio, or lubrication modes

Transition diagrams characterize critical conditions of operational parameters that separate regimes of efficient performance for a given tribosystem from regimes of inefficient performance (failure).

The third form of data presentation is the tribomap, which characterizes operational conditions that separate different wear mechanism regimes or different ranges of wear data.

Tribographs

The presentation of friction and wear data as a function of operational and structural parameters will be illustrated by some simplified examples of typical tribographs in this section. Emphasis is placed on the dependence of tribodata on time, operational parameters, structural parameters, and interaction parameters.

Dependence of Tribodata on Time

The simplest type of tribotest involves subjecting a given tribosystem to a defined constant set of structural and operational parameters and measuring friction or wear as functions of time only. The resulting observed friction or wear master curves in dry sliding are discussed in a simplified generalizing manner in the following paragraphs.

Friction-Time Master Curves. The typical friction-time curve during dry sliding consists of a simplified presentation of four stages of friction coefficient (Ref 1, 2). A typical example for a metal/metal system is shown in Fig. 1.

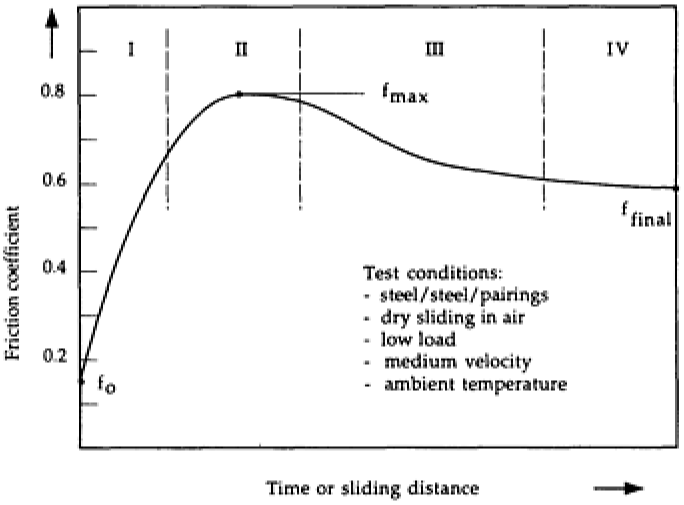


Fig. 1 A friction-time master curve

The initial value of the friction coefficient value of stage I, which is usually about f_0 ~0.1, is dependent on low loads, F_N, and on the shear resistance of surface contaminants, but is largely independent of material combinations. Surface layer removal and an increase in adhesion due to the increase in clean interfacial areas as well as increased asperity interactions and possible wear particle entrapment lead to a gradual increase in the friction coefficient.

Stage II, which produces the maximum value of the friction coefficient (f_{max} ~0.3 to 1.0 for most metal pairs), is reached when maximum interfacial adhesion, asperity deformation, and wear particle entrapment occur.

In stage III, a decrease in the friction coefficient may occur due to the possible formation of protective tribochemical surface layers and a decrease in plowing and asperity deformation processes.

Stage IV is characterized by steady-state interfacial tribological conditions leading eventually to almost constant friction coefficient values.

It should be noted that Fig. 1 shows a simplified smoothed friction graph, which in practice may be overlapped by short-term fluctuations, friction peaks, or stick-slip effects. It follows then that friction behavior should be characterized by the following (average) data in addition to the friction graph:

- Initial friction coefficient, f_0
- Maximum friction coefficient, f_{max}
- Friction coefficient at the end of test, f_{final}

Wear-Time Master Curve. The typical wear-time curve of dry sliding metal/metal systems consists of three wear stages (Ref 3). A typical example is shown in Fig. 2.

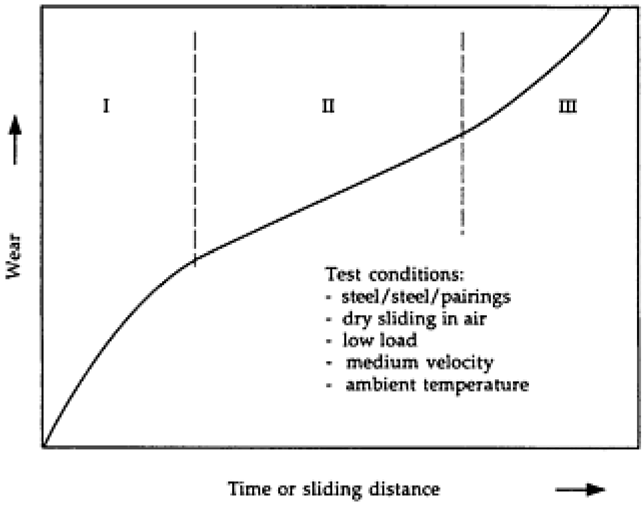


Fig. 2 A wear-time master curve

Stage I, which may have been preceded by a non-wear incubation period, is called the running-in period. During stage I, the probability of the occurrence of elementary wear events may decrease if, through changes in surface topography, the interaction rate of surface asperity collisions decreases. In the simplest case of this "self-accommodation" of the tribosystem, the amount of wear volume *W* may be given by a square-root function of time:

$$\frac{dW}{dt} \sim \frac{1}{W} \rightarrow W(t) = \text{const } t^{1/2} \quad \textbf{(Eq 1)}$$

In stage II, the tribosystem may exhibit relatively stable behavior under the action of the tribological processes. In this case the probability of wear events remains constant. This steady-state situation is characterized by a constancy of the wear-loss output per unit of time without change in the tribological processes. For engineering applications, tolerable wear coefficient values should be in the range of $k < 10^{-6}$ $mm^3/N \cdot m$. It follows that:

$$\frac{dW}{dt} = \text{const} \rightarrow W(t) = \text{const} \cdot t \quad \textbf{(Eq 2)}$$

In stage III an acceleration of wear may occur through an accumulation of elementary wear processes. During stage III, changes in the state of the system are of a directed nature and the increments in the wear processes in this regime are mutually dependent. Large increments in one interval of time cause even larger increments in the next time interval. It follows that:

$$\frac{dW}{dt} \sim W \rightarrow W(t) \sim e^{\text{const} \cdot t} \quad \textbf{(Eq 3)}$$

In this case a self-acceleration of the wear process may occur that can lead to catastrophic damage, that is, failure of the entire system.

Dependence of Tribodata on Operational Parameters

In addition to the dependence on time and motion, tribodata also depend on the set of operational parameters, which include:

- Load, F_N
- Velocity, *v*, in sliding, rolling, and so forth
- Temperature, *T*, at stated location and time

Therefore, in the presentation of tribodata, the primary influencing operational parameters must be carefully identified as will be illustrated in the following examples of tribographs for steels, ceramics, and polymers.

Dependence of Wear on Load and Velocity. A well-known characteristic example of the dependence of wear on load, F_N, and sliding velocity, *v*, is shown in Fig. 3 (Ref 4). Gradual or sharp transitions between mild or severe wear are caused by a change in the structure of oxides (α-Fe_2O_3, Fe_3O_4, or FeO) on the sliding surfaces or by the breaking away of protective oxide films (Ref 5, 6). Above these transitions, wear increases about linearly (on a logarithmic scale) with increasing load up to a second transition, which is recognized by a sharp drop in wear. This is due to the formation of a new protective oxide film that forms at high-contact temperature (Ref 7).

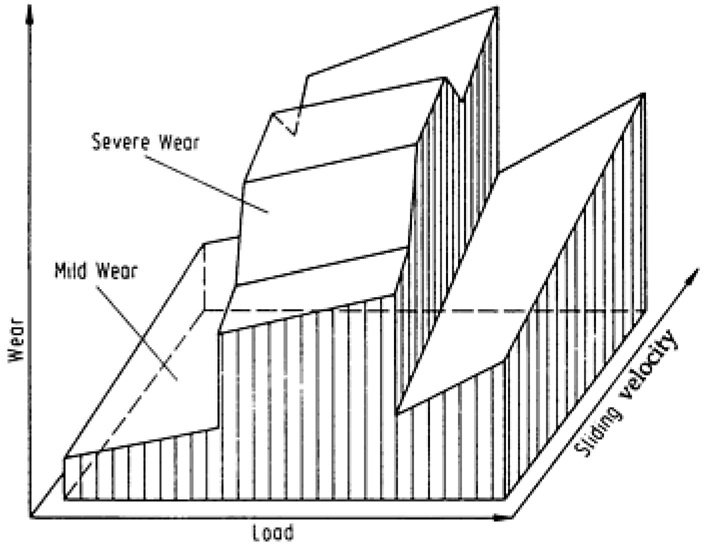


Fig. 3 Wear intensity of steels in dry sliding as a function of load and sliding velocity (logarithmic scales)

Dependence of Wear on Velocity and Temperature. The dependence of the wear coefficient on sliding velocity, *v*, and bulk specimen temperature, *T*, is illustrated for dry-sliding ceramic pairs (partially stabilized zirconia, MgO-PSZ) in Fig. 4, which shows a gradual changeover from low wear to high wear (Ref 8). As functions of the increase of *v* and *T*, the initial monoclinic and tetragonal ZrO_2 phases change under the high hot-spot temperatures and microasperity pressures to cubic ZrO_2, with a volume reduction of up to 9 vol%. This leads to increased wear due to surface tension and then to intergranular microcracking.

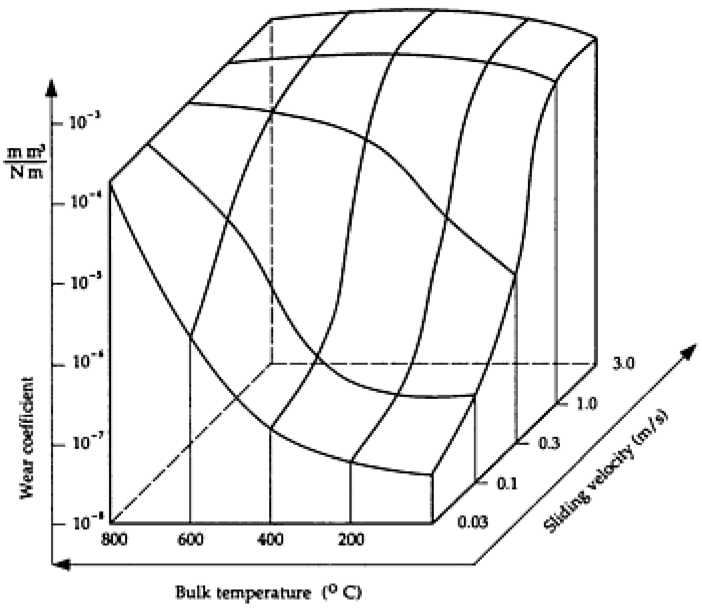


Fig. 4 Wear coefficient of zirconia ceramics in dry sliding as a function of sliding velocity and temperature

Dependence of Friction on Load, Velocity, and Temperature. The dependence of the friction coefficient on the set of operational parameters (F_N,v,T) for a polymer/steel sliding system was described in the previous article in this Section on "Design of Friction and Wear Experiments." Figure 4 for the same article shows the relationship of the friction coefficient with F_N, v, and T.

Dependence of Tribodata on Structural Parameters

Tribodata are also closely related to the structural components of tribosystems. Key structural parameters include:

- Composition
- Roughness
- Modulus
- Hardness
- Toughness
- Microstructure

The dependence of tribodata on structural parameters, which was described in detail in the article "Basic Tribological Parameters" in this Section, will be illustrated in this article by considering only the example of the abrasive wear resistance of material (Fig. 5). For metals, the increment in abrasion resistance with increasing hardness is substantially larger in pure metals than in heat-treated steels. At the same bulk hardness, steels with higher carbon content show higher abrasion resistance. Ceramics show lower abrasion resistance than pure metal or annealed steels of comparable hardness when hard abrasives are used. Polymers exhibit low abrasion resistance due to their low hardness. For thermoplastics, the abrasive wear resistance depends also on another important structural parameter, namely the rupture strength (Ref 9), as can be seen from Fig. 6.

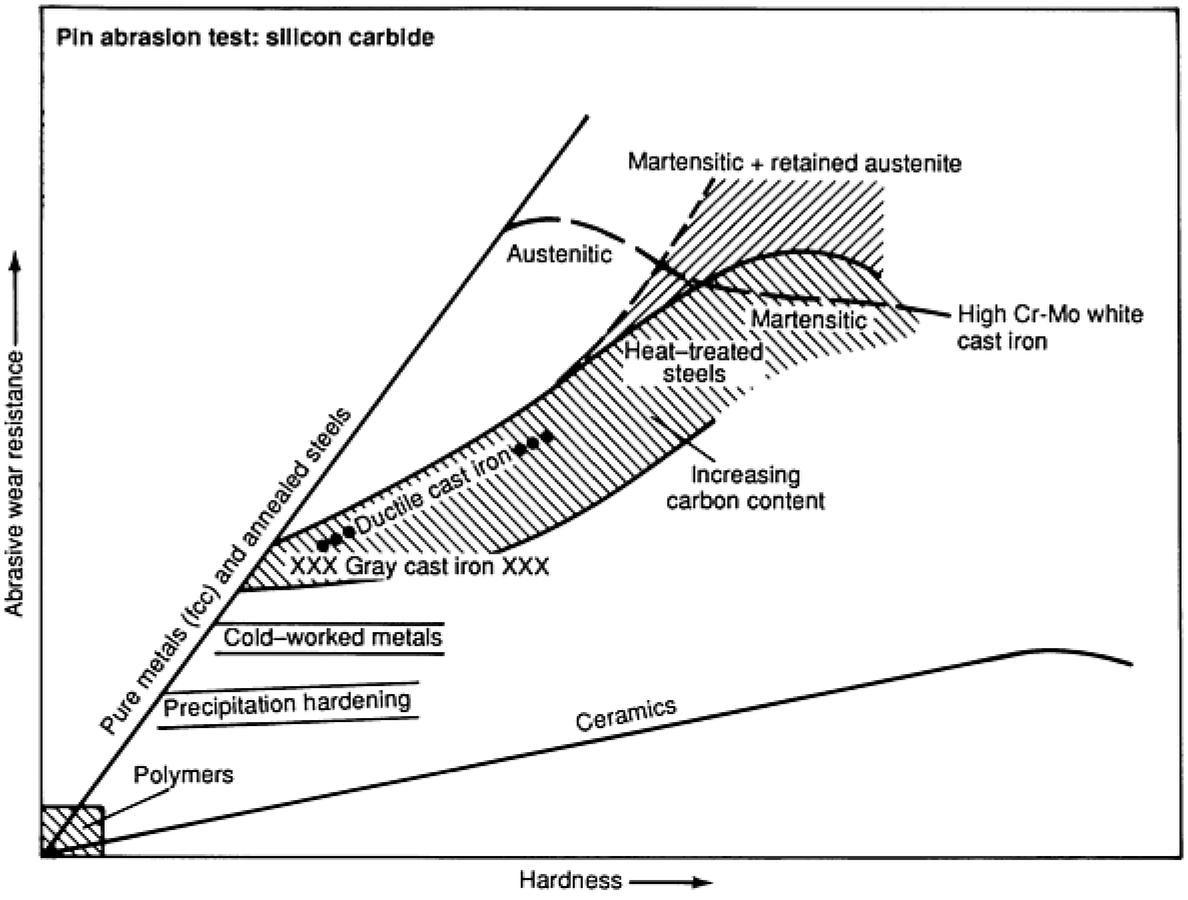


Fig. 5 Abrasive wear resistance of different materials as a function of bulk hardness

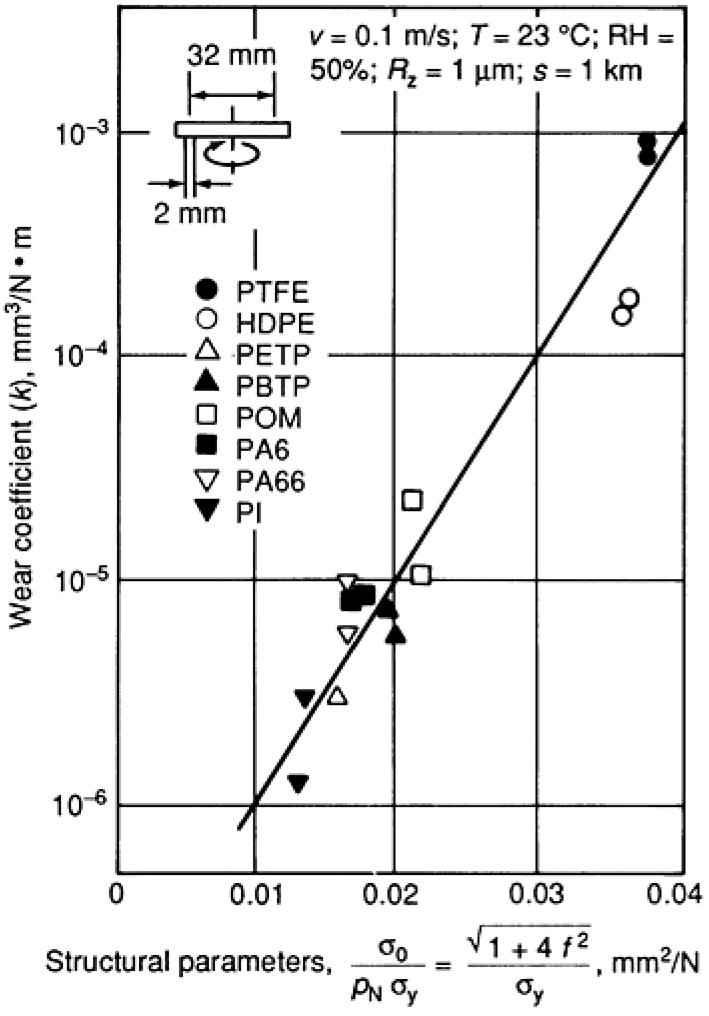


Fig. 6 Abrasive wear coefficient of polymer/steel (AISI 52100 steel, 800 HV, roughness R_z = 1 μm) sliding pairs as a function of stress and strength parameters. p_N, contact pressure; σ_0, interfacial stress; σ_y, polymer rupture strength; *f*, friction coefficient. Source: Ref 9

Finally, it should also be emphasized that for an unambiguous presentation of wear data, a distinction must be made between system wear and component wear. As can be seen from Fig. 7, for a Si_3N_4 ceramic, different values for component wear (ball or disk specimen) and the system wear of both partners of a ball-on-disk configuration result.

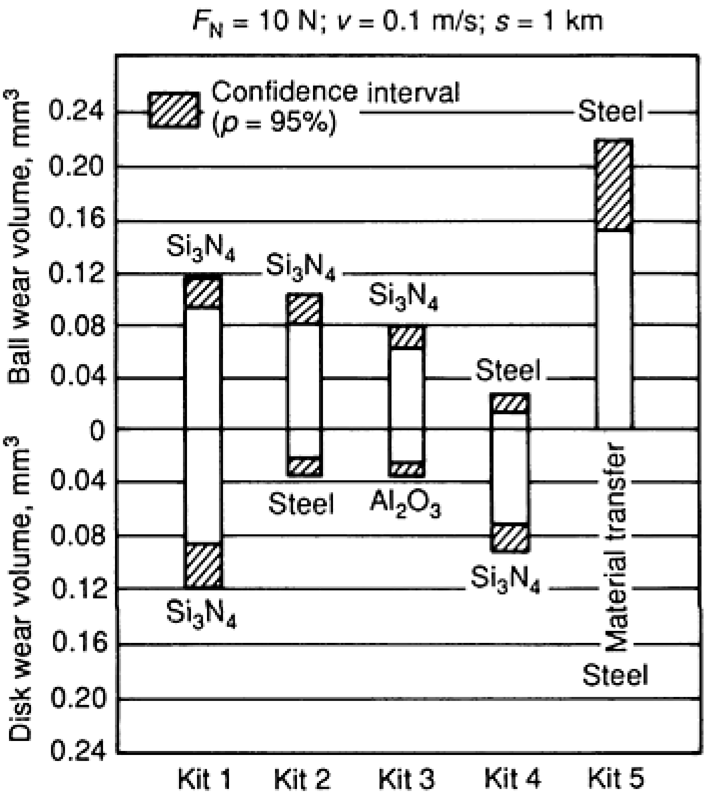


Fig. 7 Presentation of wear data for the components of a laboratory tribotesting system (ball-on-disk configuration, VAMAS reference conditions as described in the article "Design of Friction and Wear Experiments")

Dependence of Tribodata on Interaction Parameters

The dependence of tribodata on interaction parameters is the key feature of friction and wear data as compared with other materials-related quantities, such as strength data. Interaction parameters are in existence only during the operation of the tribosystem under consideration. They determine the dynamic state of the tribosystem and are not intrinsically derivable from the static individual tribocomponents. As discussed in the article "Basic Tribological Parameters" in this Section, fundamental interaction parameters that influence friction and wear data include:

- Interface forces and energies (connected with interface physics and chemistry)
- Contact deformation modes
- Contact stresses
- Contact area/wear-track ratio
- Lubrication modes

Furthermore, interaction parameters must characterize possible interfacial "third-body processes" (Ref 10), which involve the compositional, microgeometrical, and microstructural changes of interfaces and the generation, formation, and detachment of wear particles.

Transition Diagrams

Regimes of efficient and inefficient performance of tribosystems may be distinguished by transition diagrams. For lubricated sliding Hertzian (steel) contacts, a methodology for the determination of transition diagrams has been developed by the International Research Group on Wear of Engineering Materials (IRG-OECD) (Ref 11, 12). The primary features of this methodology are summarized below.

Step 1. Tribological tests should be run with a defined lubricated Hertzian contact system (for example, a ball-on-ring or ball-on-flat specimen configuration) under constant operating conditions of sliding velocity, *v*, bulk oil temperature, *T*, sliding distance, *s*, and test duration, *t*, with a stepwise increased load, F_N. The tests may be run either with new specimens for each load step (that is, a "no run-in procedure" where $t < 10$ s), or with only one specimens set (that is, a "run-in procedure"). The friction coefficient, *f*, and wear volume *W* are then measured and the wear coefficient, $k = W/(F_N \cdot s)$, is calculated.

Step 2. For a constant set of (*v*,*T*,*t*,*s*), the critical load F_{Ncrit} for the transition from regime I (partial elastohydrodynamic, EHD, lubrication) to regime II (incipient scuffing), as shown in Fig. 8, is characterized by:

$$f < 0.1 \rightarrow f > 0.3$$
$$k < 10^{-6}\ \text{mm}^3/\text{N} \cdot \text{m} \rightarrow k > 5 \times 10^{-6}\ \text{mm}^3/\text{N} \cdot \text{m} \qquad \textbf{(Eq 4)}$$

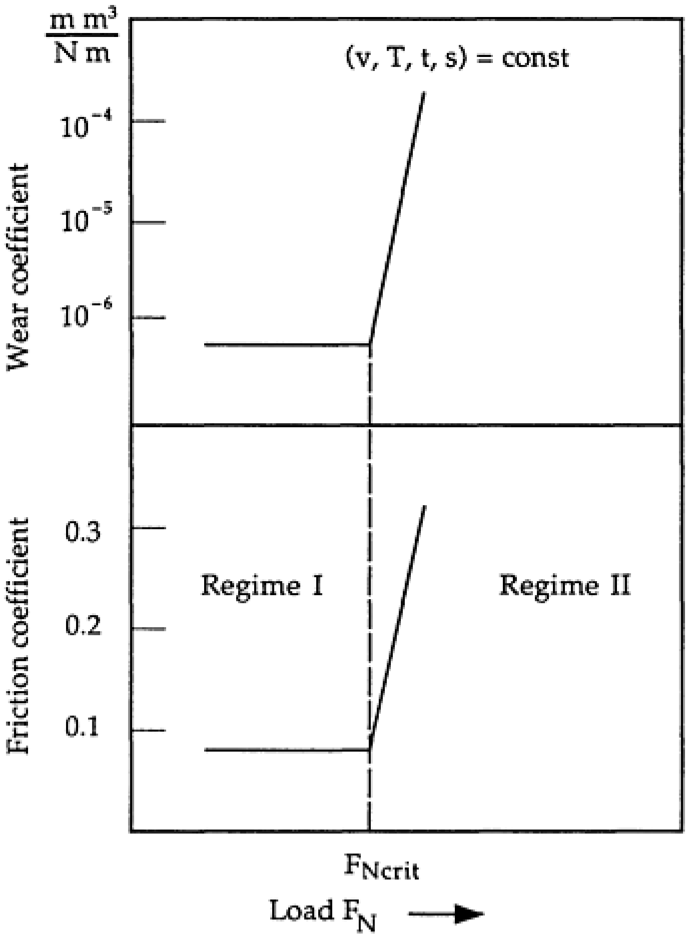


Fig. 8 Determination of critical failure loads, F_{Ncrit}, for sliding lubricated concentrated contacts

During step 3, the critical load, F_{Ncrit} for a variation of sliding velocity, *v* (the other parameters *T*, *t*, and *s* being held constant), is determined. The graph of F_{Ncrit} versus the corresponding values of *v* is called the IRG transition diagram (Fig. 9). Note that in addition to the I/II transition, there is a II/III transition that characterizes the change from incipient scuffing to complete seizure.

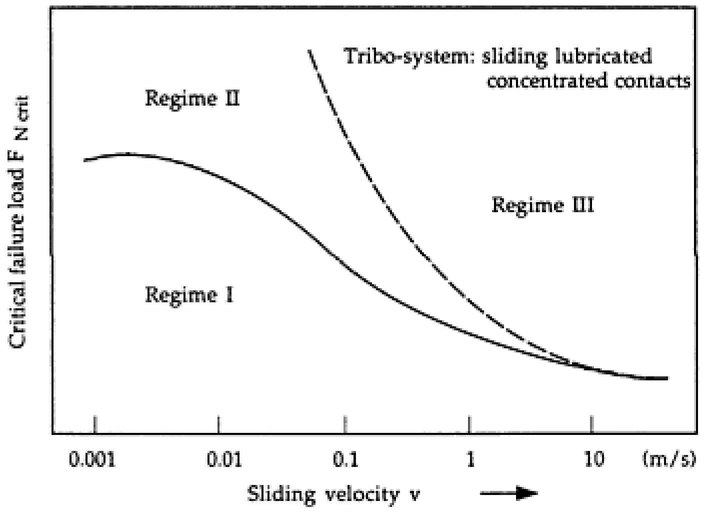


Fig. 9 IRG transition diagram for sliding lubricated concentrated contacts (critical failure load as a function of sliding velocity)

A F_{Ncrit}/T transition diagram characterizing the critical failure load as a function of the bulk oil temperature, *T*, (for constant *v*,*t*,*s*) may also be determined. In combining these two diagrams, a three-dimensional failure surface results (Ref 13), which characterizes the load-carrying capacity of lubricated concentrated contacts as a function of sliding velocity, *v*, and bulk oil temperature, *T* (Fig. 10).

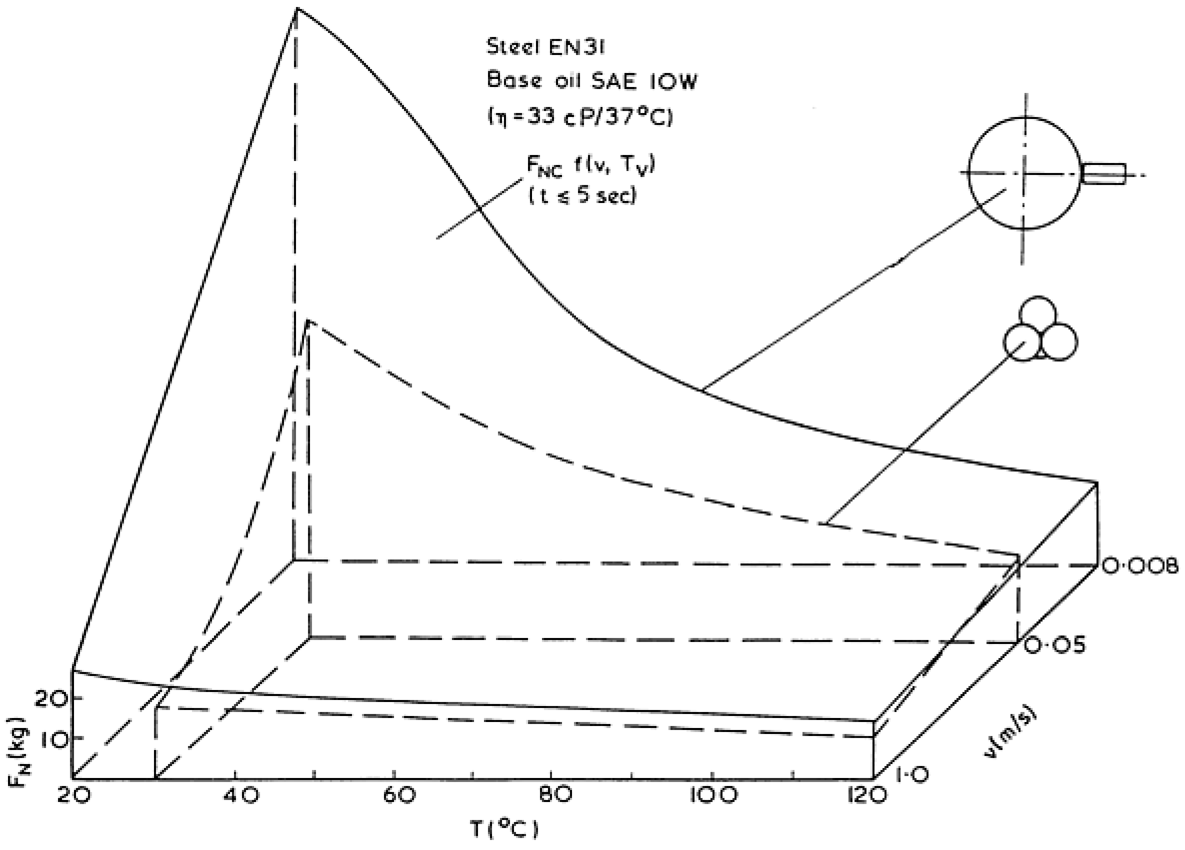


Fig. 10 Critical failure load , F_{Ncrit}, of sliding lubricated concentrated contacts as a function of sliding velocity, *v*, and bulk oil temperature, *T*. Oil viscosity (given in cP) is indicated by η.

With the use of IRG transition diagrams, the influences of materials properties, lubricant characteristics (for example, viscosity or chemical additives), or environmental conditions (for example, inert gas or humidity) on the functional limits of the operational variables of sliding lubricated contacts can be characterized (Ref 11, 12, 13).

Tribomaps

Tribomaps characterize regimes of different tribological processes (or different ranges of friction and wear data) by giving boundaries of operational parameters for these regimes. For the design of empirical wear mechanism maps, Ashby and co-workers (Ref 14, 15) have used normalized values for wear rate, load, and sliding velocity defined by:

$$\tilde{W} = W/A_n$$
$$\tilde{F}_N = F_N/A_n \cdot H_0 \qquad \textbf{(Eq 5)}$$
$$\tilde{v} = v \cdot r_0/a$$

where A_n is the nominal (apparent) contact area of the wearing surface, H_0 is the room-temperature hardness, a is the thermal diffusivity, r_0 is the radius of the circular nominal contact area, $\tilde{W}$ is the volume lost per unit area of surface, per unit distance slid; $\tilde{F}_N$ is the nominal pressure divided by the surface hardness, and $\tilde{v}$ is the sliding velocity divided by the velocity of heat flow.

In analyzing the results of wear tests published in the literature, the different dominating wear mechanisms were classified into the following categories:

- Seizure
- Melt-dominated wear
- Oxidation-dominated wear (mild and severe oxidational wear)
- Plasticity-dominated wear (including delamination wear)

For each wear regime, the boundaries of the operational parameters F_N and v were estimated; the resulting wear map is shown in Fig. 11. Additional information on wear maps can be found in the article "Friction and Wear of Cutting Tools and Cutting Tool Materials" in this Volume.

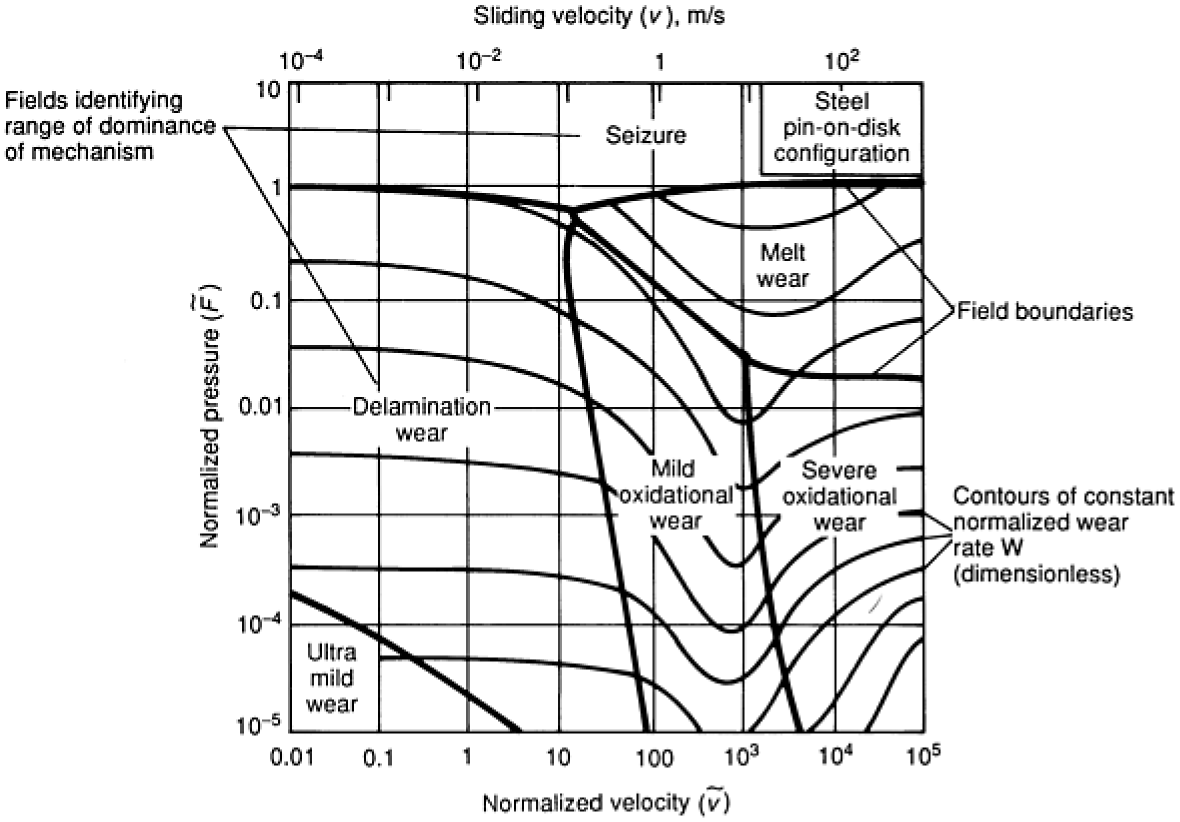


Fig. 11 Empirical wear mechanism map for steel (pin-on-disk configuration)

References

1. N.P. Suh, *Tribophysics,* Prentice-Hall, 1986, p 73
2. P.J. Blau, *Friction and Wear Transitions of Materials,* Noyes Publications, 1989
3. H. Czichos, *Tribology--A Systems Approach to the Lubrication and Wear,* Elsevier, Amsterdam, 1978, p 195, 196
4. K.-H. Zum Gahr, *Microstructure and Wear of Materials,* Elsevier, Amsterdam, 1987, p 379
5. N.C. Welsh, The Dry Wear of Steels, *Philos. Trans. R. Soc. (London) A,* Vol 257, 1965, p 31-70
6. K.-H. Habig, *Wear and Hardness of Materials,* Hanser-Verlag, München, 1980 (in German)
7. T.F.J. Quinn, D.M. Rowson, and J.L. Sullivan, Applications of the Oxidational Theory of Mild Wear to the Sliding Wear of Low Alloy Steel, *Wear,* Vol 65, 1980, p 1-20
8. M. Woydt, D. Klaffke, K.-H. Habig, and H. Czichos, Tribological Transition Phenomena of Ceramic Materials, *Wear,* Vol 136, 1990, p 373-380
9. H. Czichos, Influence of Adhesive and Abrasive Mechanisms on the Tribological Behaviour of Thermoplastic Polymers, *Wear,* Vol 88, 1983, p 27-43
10. M. Godet, Third Bodies in Tribology, *Proceedings of EUROTRIB 1989,* K. Holmberg and I. Nieminen, Ed., The Finnish Society for Tribology, ESPO, Vol 1, 1989, p 1-15
11. H. Czichos and K. Kirschke, Investigations into Film Failure (Transition Point) of Lubricated Concentrated Contacts, *Wears,* Vol 22, 1972, p 321-336
12. C.M. Lossie, J.W.M. Mens, and A.W.J. de Gee, Practical Applications of the IRG Transition Diagram Technique, *Wear,* Vol 129, 1989, p 173-182
13. H. Czichos, Failure Criteria in Thin Film Lubrication: The Concept of a Failure Surface, *Tribol. Intl.,* Vol 7, 1974, p 14-20
14. S.C. Lim and M.F. Ashby, Wear-Mechanism Maps, *Acta Metall.,* Vol 35, 1987, p 1-24

15. S.C. Lim, M.F. Ashby, and J.H. Bruton, Wear-Rate Transitions and their Relationship to Wear Mechanisms, *Acta Metall.*, Vol 35, 1987, p 1343-1348

Concepts of Reliability and Wear: Failure Modes

Horst Czichos, BAM (Germany)

Introduction

METHODS TO CHARACTERIZE THE RELIABILITY of mechanical equipment on the basis of measurements or estimations of wear are often tried in mechanical engineering applications (Ref 1). In this article, statistical techniques and probability concepts for the evaluation and presentation of wear and reliability data are briefly outlined. More detailed information on reliability analysis can be found in the references provided as well as Vol 17 of *ASM Handbook* (formerly 9th Edition *Metals Handbook*).

Acknowledgements

The author would like to thank his tribology colleagues at BAM, and in particular Karl-Heinz Habig and Erich Santner for their help in the preparation of this Section. Special thanks are also due to BAM mathematicians. Wolfgang Gerisch and Thomas Fritz for their review and valuable contributions.

Characteristics of Reliability

Reliability is defined as "the probability of a device performing its purpose adequately for the period of time intended under the operating conditions encountered." (This is the classic definition of reliability given by the Radio Electronics and Television Manufacturers Association in 1955.) In a quantitative way, the reliability of systems, devices, or products may be characterized by the probability concepts outlined in Table 1.

Table 1 Reliability probability concepts

Equation	Concept
$F(t)$	Probability distribution function of the lifetime random variable
$f(t) = \frac{dF(t)}{dt}$	Density function
$\lambda(t) = \frac{f(t)}{1 - F(t)}$	Failure rate
$R(t) = 1 - F(t) = \exp\left(-\int_0^t \lambda(\tau)d\tau\right)$	Reliability function: probability that the system or device will not fail during the time period $(0, t)$
$\text{MTTF} = \int_0^\infty t f(t)dt$	Mean time to failure

Relationship Between Wear and Reliability

Consider as a starting point for the reliability considerations of tribosystems that wear behavior is a function of time (Ref 2). In the previous article on "Presentation of Friction and Wear Data," it was explained that for time-dependent wear of a tribosystem, three different wear stages can be distinguished:

- Self-accommodation or running-in wear
- Steady-state wear
- Self-acceleration (catastrophic damage) of wear

These wear mode changes in the system behavior may follow each other with time, as indicated in Fig. 1. In this figure, W_{lim} denotes a maximum admissible level of wear losses. At this level the system structure has changed in such a way that the functional input-output relations of the system are disturbed severely. Repeated measurements show random variations in the data, as indicated by the dashed lines in Fig. 1.

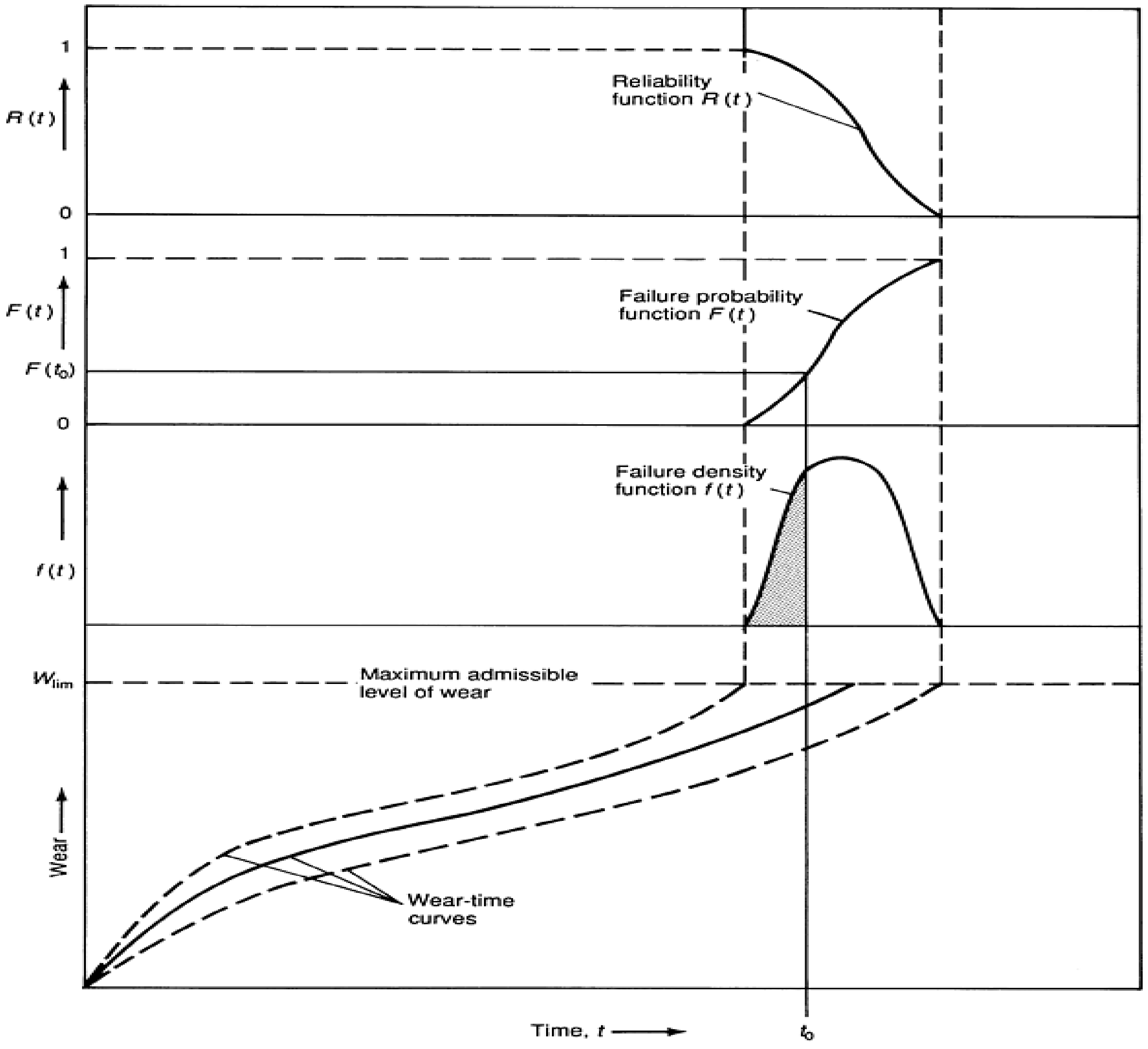


Fig. 1 Simple curves of wear and failure and reliability functions

From sample curves of wear, a probability density function, $f(t)$, of the time for reaching the maximum admissible level of wear (W_{lim} = constant) is obtained. For a given time, t_0, the shaded area under the curve $f(t)$, that is, the value of $F(t_0)$, is a measure of the probability that the system fails within the time $t < t_0$.

Statistical Distributions of Wear and Reliability

For the modeling of the distribution of measured wear data and the estimation of reliability data, statistical distributions may be used. In the following paragraphs, some of the statistical distributions that have been applied for these purposes are briefly reviewed in a highly simplified manner in order to illustrate the given probability concepts. With respect to the estimation of the occurring parameters, the calculation of confidence intervals, and related subjects, the reader is referred to Ref 3, 4, 5, 6.

For exponential distribution:

$$\begin{aligned} \lambda(t) &= \lambda = \text{const} \\ f(t) &= \lambda \cdot \exp(-\lambda t) \\ R(t) &= \exp(-\lambda t) \\ \text{MTTF} &= 1/\lambda \end{aligned} \qquad \textbf{(Eq 1)}$$

In this case, the failure rate is constant, which means that failure occurs accidentally without an accumulation of fatiguelike effects during service time. Components in a machine fail in this mode, for example, when the failure is brittle fracture. In Fig. 2 the density function of the failure of a diesel engine control unit is plotted showing an exponential distribution (Ref 7).

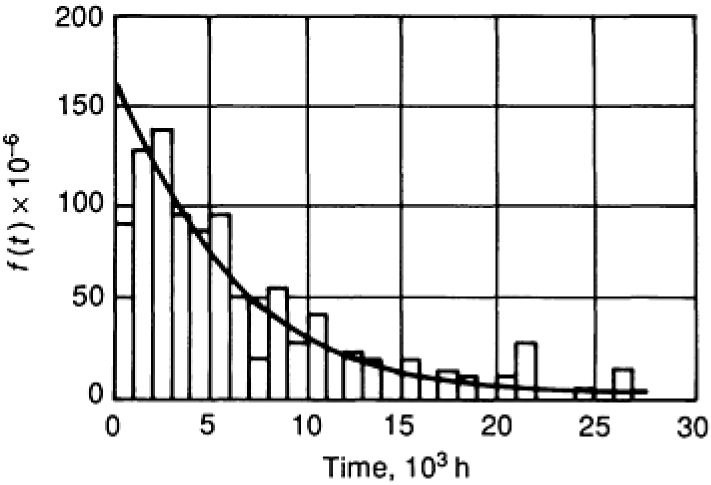


Fig. 2 Failure density function of diesel engine control units

For normal distribution:

$$f(t) = \left(1/\sigma\sqrt{2\pi}\right)\exp\left\{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right\}$$

$$\lambda(t) = \frac{f(t)}{1-\phi((t-\mu)/\sigma)} \qquad \textbf{(Eq 2a)}$$

where ϕ is the standard normal distribution function, and

$$\begin{aligned} R(t) &= 1 - \phi((t-\mu)/\sigma) \\ \text{MTTF} &= \mu \end{aligned} \qquad \textbf{(Eq 2b)}$$

Many machine components obey this distribution, especially if the failure occurs due to wear processes.

For lognormal distribution:

$$f(t) = \left(1/t\sigma\sqrt{2\pi}\right)\exp\left\{-\frac{1}{2}\left(\frac{\ln(t)-\mu}{\sigma}\right)^2\right\}$$

$$R(t) = 1 - \phi\left(\frac{\ln(t)-\mu}{\sigma}\right) \quad \text{(Eq 3)}$$

$$\lambda(t) = \frac{f(t)}{1-\phi((\ln(t)-\mu)/\sigma}$$

$$\text{MTTF} = \exp\,(\mu + \sigma^2/2)$$

This distribution is concentrated on the positive t-axis. Its failure rate increases to a maximum and then decreases to zero. Therefore, it can be used for modeling survival times after extreme stress.

For the Weibull distribution:

$$\lambda(t) = \frac{C}{t_0}t^{c-1}$$

$$f(t) = \frac{C}{t_0}t^{c-1}\exp\,(-t^c/t_0) \quad \text{(Eq 4)}$$

$$R(t) = \exp(-t^c/t_0)$$

$$\text{MTTF} = t_0^{1/c}\,\Gamma((C+1)/C)$$

In its simplest form, this is a distribution with two key parameters: t_0, the nominal life, and the constant C. This distribution is found to represent failure of many kinds of mechanical systems, such as fatigue in ball bearings. As an example, Fig. 3 shows the probability distribution function of the time to failure, $F(t)$, as determined by testing 500 grease-lubricated ball bearings at 1000 rpm (Ref 8).

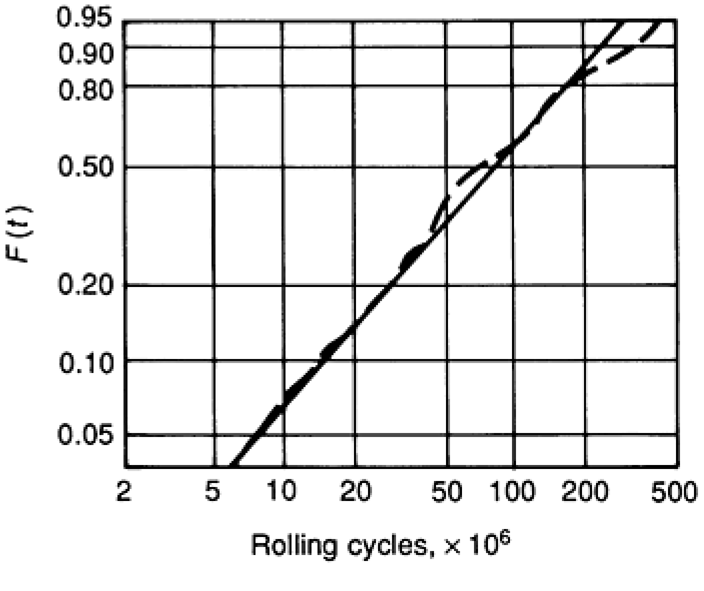


Fig. 3 Failure distribution function of ball bearings

For the Gamma distribution:

$$f(t) = C \frac{(Ct)^{x-1}}{\Gamma(x)} \exp(-Ct)$$

$$\lambda(t) = C\, t^{x-1} \exp(-Ct)/[\Gamma(x) - \Gamma(x, Ct)]$$

(Eq 5a)

where $\Gamma(a,z)$ is the standard incomplete gamma function, and

$$R(t) = (\Gamma(x) - \Gamma(x,Ct))/\Gamma(x)$$
$$\text{MTTF} = x/\text{C}$$

(Eq 5b)

where $\Gamma(x)$ is a Gamma function. This is also a distribution with two parameters. Theoretically, the importance of this distribution is attributed to the fact that the equation is an *x*-fold convolution of the exponential function. It means physically that a component fails at *x*th shock which occurs as a Poisson statistical process with parameter C. In Fig. 4, the density function of the failure of a piston in a diesel engine is plotted showing a Gamma distribution with *x* = 2 (Ref 7).

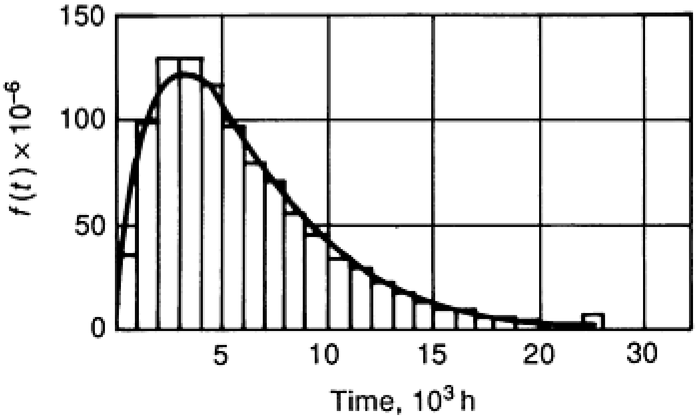


Fig. 4 Failure density function of diesel engine pistons

Wear and Failure Modes

The aforementioned examples of statistical distributions indicate that different failure modes and different elementary failure processes are associated with different types of failure distribution functions. It follows that from the experimental determination of failure distribution curves, conclusions may be drawn as to the type of failure mechanism. For most tribosystems failing as a consequence of wear processes, the failure behavior is characterized by the normal distribution or the Weibull distribution. If for a given type of tribosystem the failure mode and the type of failure distribution are known, this knowledge can be used to improve the reliability of the system (Ref 9). For instance, this approach can be used to select the type of a ball or roller bearing system to operate under a given set of operating conditions with high operational safety (Ref 10). More general treatments concerning correlations between life data, statistical lifetime distribution, and failure models can be found in Ref 3, 4, 5, 6, 11 and 12.

Dependence of Failure Rate on Operating Duration

To conclude the brief discussion on the failure and reliability of tribosystems, and dependence of the failure rate on the operating duration of a system should be considered. If the failure rate is plotted as a function of time, a curve known as the "bathtub curve" is generated (Fig. 5).

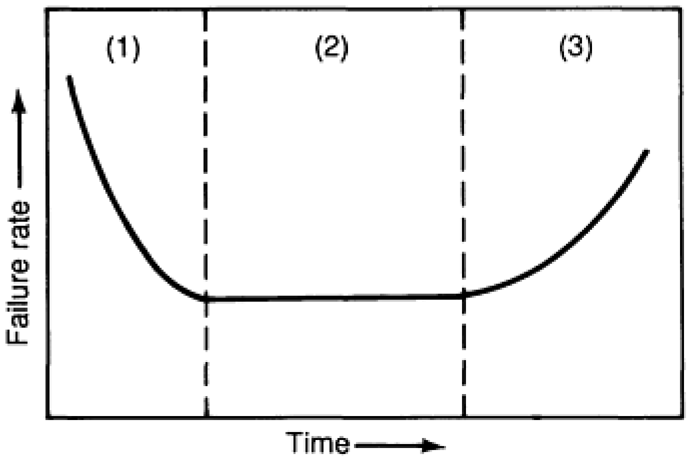


Fig. 5 "Bathtub" failure rate curve

In this curve, three regimes can be distinguished: (1) early failures, (2) random failures, and (3) wear-out failures. None of the distribution curves discussed above have this bathtub-shaped failure curve, but an approximation can be obtained by selecting an appropriate probability density function for each of the three regimes. Regime (1) describes the region of the "infant death" of the system. This regime is characterized by a decrease of the failure rate with time during running-in wear. Regime (2), which exhibits a constant failure rate, is the region of normal running. Here failure occurs as a consequence of statistically independent factors. Regime (3) is characterized by an increase of the failure rate with time. Here failure may be due to aging effects. As described above, for a great deal of tribo-induced failures, the failure rate increases with time. Thus region (3) of the bathtub curve of Fig. 5 appears to be relevant for the normal mode of wear-induced failure of mechanical systems.

References

1. G. Fleischer, H. Gröger, and H. Thum, *Wear and Reliability,* VEB Verlag Technik, Berlin, 1980 (in German)
2. H. Czichos, *Tribology--A Systems Approach to the Science and Technology of Friction, Lubrication and Wear,* Elsevier, Amsterdam, 1978, p 234-240
3. W. Nelson, *Applied Life Data Analysis,* John Wiley, 1982
4. J.F. Lawless, *Statistical Models and Methods for Lifetime Data,* John Wiley, 1982
5. N.R. Mann, R.E. Schafer, and N.D. Singpurwalla, *Methods for Statistical Analysis of Reliability and Life Data,* John Wiley, 1974
6. J. McCormick, *Reliability and Risk Analysis,* Academic Press, 1981
7. G. Fleischer, Problems of the Reliability of Machines, *Wiss. Z. Tech. Univ. Magdeburg,* Vol 16, 1972, p 289 (in German)
8. G. Bergling, Reliability of Rolling Bearings, *Kugellager-Zeitschrift,* Vol 51, 1976, p 1 (in German)
9. A. Holfeld, Wear and Life Time Determination on a Statistical Basis, *Schmierungstechnik,* Vol 20, 1989, p 167-171 (in German)
10. A. Sturm, "Rolling Bearing Diagnosis in Machines and Plants," Verlag TÜV Rheinland, Cologne, 1986 (in German)
11. H.F. Martz and R.A. Walter, *Bayesian Reliability Analysis,* John Wiley, 1982
12. A.E. Green and A.J. Bourne, *Reliability Technology,* Wiley-Interscience, London, 1972

Friction and Wear of Rolling-Element Bearings

Tedric A. Harris, Pennsylvania State University

Introduction

ROLLING-ELEMENT BEARINGS, also called rolling bearings and antifriction bearings, are among the most common machine elements. The basic design of the current ball bearings was devised by Leonardo da Vinci in the 15th century. (See Ref 1 for a history of lubrication and bearings.) The term *rolling element* refers to the ball or roller components that are used to separate the inner and outer rings. The term *antifriction* is used because the bearings tend to have very low friction characteristics compared to fluid film bearings or simple sliding bearings. In addition to the components cited above (that is, balls or rollers, inner rings, and outer rings), most rolling bearing have a *cage* (also called a *separator* or *retainer*) that spaces the rolling elements during operation. The cage also serves to retain the rolling elements in the bearing prior to assembly and also in the subsequent application. Figure 1 illustrates the components in a typical ball bearing. Note that the rolling elements run on an inner ring track called the inner raceway; similarly, the outer ring track is called the outer raceway.

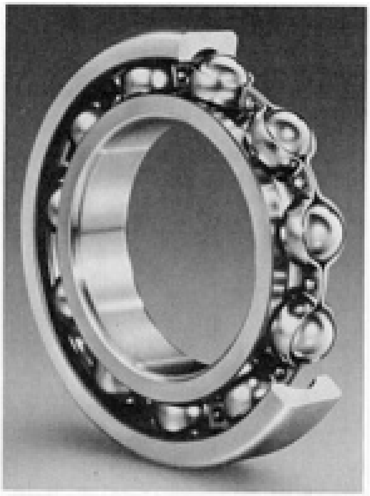


Fig. 1 Cutaway view of radial ball bearing showing inner ring, outer ring, balls, and cage assembly

Rolling bearings have much less friction torque than conventional hydrodynamic bearing types, and starting friction torque is only marginally greater than operating friction torque. In addition, rolling-element bearing deflection is not as sensitive to load fluctuation as is deflection in a hydrodynamic bearing. In most applications, only a small quantity of lubricant is required for satisfactory operation, eliminating the need for expensive and space-consuming lubricating systems. Moreover, rolling-element bearings require less space than corresponding hydrodynamic bearings, and they can be selected or designed in compact units to support combination loads (for example, radial, thrust, and moment loads). The load and speed ranges to which a given rolling-element bearing may be subjected and function efficiently under each condition are significantly wide. In general, if a rolling-element bearing can satisfy the operating conditions for a given application, it represents the economical choice.

Types of Rolling-Element Bearing

There are two basic rolling-element types: balls and rollers. Ball bearings enjoy the most universal usage. However, for applications in which very heavy loads must be supported, roller bearings find extensive usage. In addition to categorizing

rolling bearings according to rolling-element type, such bearings are also identified according to the predominant loading they are designed to support (for example, radial load or thrust load [also called axial load]). Within each subcategory type (for example, radial ball bearings, thrust ball bearings, radial roller bearings, and thrust roller bearings), there are several different basic variations, the usage of each depending on the load and speed conditions to be accommodated.

Ball Bearings

Ball bearings can be further classified as either radial ball bearings, angular-contact ball bearings, or thrust ball bearings.

Radial Ball Bearings. The most common ball bearing type is the nonseparable Conrad assembly (Fig. 2). The nominal contact angle between a ball and a raceway is 0° (the contact angle is defined as the angle the ball-raceway load vector makes with the bearing radial plane). The bearing is designed to carry moderate radial road. The bearing can, however, support some thrust (axial) and moment load in addition to the radial load. In this case, the contact angles between balls and raceways increase beyond 0° with the application of thrust. In most cases, the bearings can operate at high speeds of rotation ($\leq 5 \times 10^5\ n \cdot d_m$, where n is the speed of rotation in rev/min and d_m is the pitch diameter, above which special cooling is often required to keep operating temperatures at <170 °C, or 338 °F). Because the bearing balls and rings form an inseparable unit when assembled, cages are either two-piece riveted assemblies or one-piece snap-on plastic (generally fiberglass-filled polyamide [nylon]) units. When using bearings with plastic cages, care must be exercised to ensure the compatibility of the cage material with the bearing lubricant and operating temperatures. The latter must not be >120 °C (>248 °F).

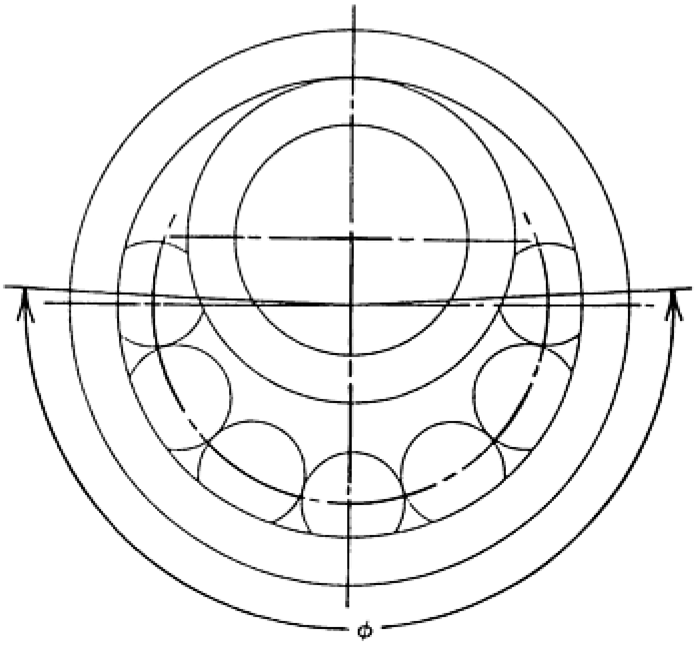


Fig. 2 Schematic showing assembly process for a nonseparable Conrad-type ball bearing. ϕ, assembly angle

Self-aligning ball bearings are principally double-row radial bearings that can accommodate radial load simultaneously with substantial misalignment (for example, from 1.5 to 3° depending on internal design), without detriment to bearing endurance. In conjunction with the principal radial load, they can also support some axial loading. Because the outer raceway is a portion of a sphere (Fig. 3), the conformity of the outer raceway to the ball is not close. Accordingly, the outer raceway has less load-carrying capacity than does the inner raceway. The reverse is true for almost all other basic types of radial rolling-element bearings.

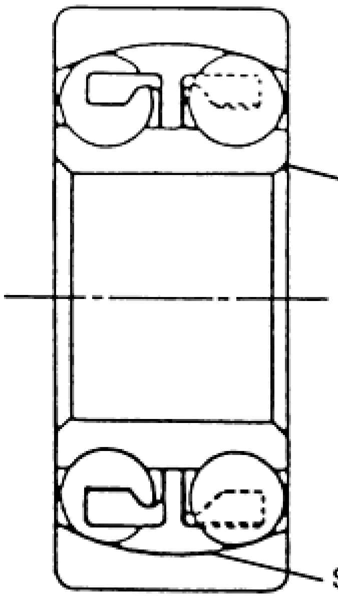


Fig. 3 Typical construction of a self-aligning ball bearing incorporating a spherically shaped outer raceway

Angular-contact ball bearings are radial bearings designed to support substantial amounts of axial load in addition to radial load. Basic catalog bearing designs have nominal contact angles from 15 to 40° (Fig. 4); the bearings having the higher contact angles can carry the greater amount of axial loading. For the 40° contact angle bearing, applied thrust load must be at least equal to radial load for satisfactory operation. The bearings can also support simple axial loading. The bearings are always mounted in pairs to accommodate thrust loading induced by applied radial loading; they may be mounted in back-to-back or face-to-face duplex arrangements (Fig. 5). The former provides substantial resistance to moment loading (that is, high stiffness), whereas the latter provides greater ability to accommodate misalignment. In addition, angular-contact bearings can be mounted in tandem for increased axial load-carrying capacity (Fig. 6). Duplex angular-contact ball bearings can be axially preloaded to achieve increased stiffness (that is, reduced deflection under applied loading).

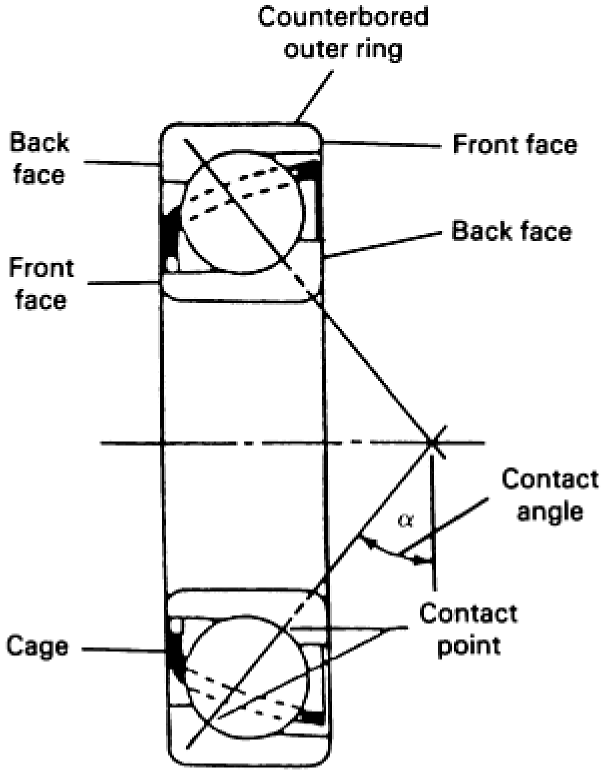


Fig. 4 Key components of an angular-contact ball bearing

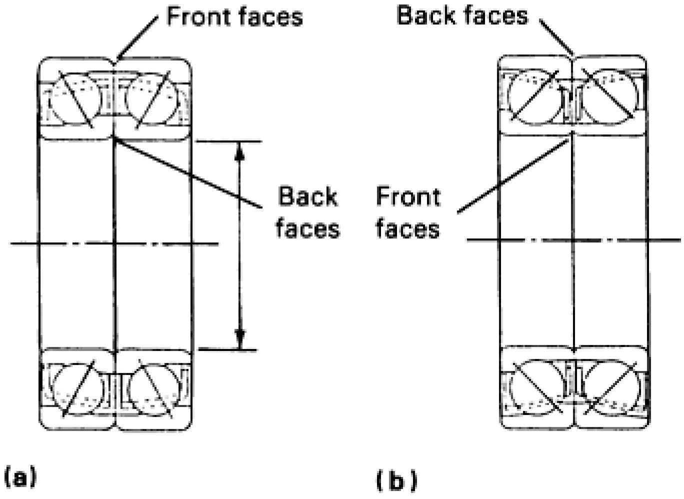


Fig. 5 Duplex angular-contact ball bearing arrangements used to accommodate thrust loading induced by applied radial loading. (a) Face-to-face mounting. (b) Back-to-back mounting

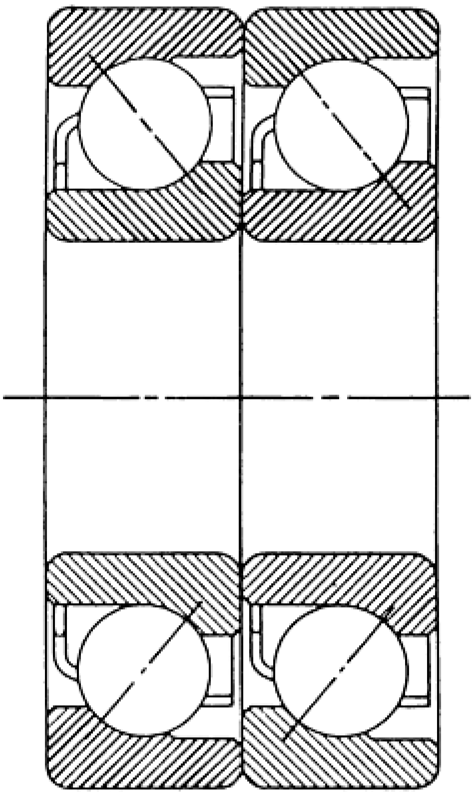


Fig. 6 Pair of angular-contact ball bearings mounted in tandem

The bearings may be assembled with one-piece pressed (stamped) or precision-machined cages fabricated from steel, brass, or bronze. Bearings having the machined cages are capable of higher speeds. Bearings can also be supplied with plastic (mainly polyamide 6/6, nylon) cages; these also have high-speed capability. The operating speed capability of angular contact ball bearings is somewhat less than that of other radial ball bearings.

Split inner ring ball bearings (Fig. 7) are double-direction angular-contact ball bearings (that is, they can support thrust load in either axial direction). When compared to duplex angular-contact bearings, however, they are narrower in width. Radial load can be supported only in conjunction with thrust loading sufficient to load all balls. Generally, the balls are spaced apart by a one-piece cage precision machined from steel, brass, or bronze. The bearings are capable of high-speed operation.

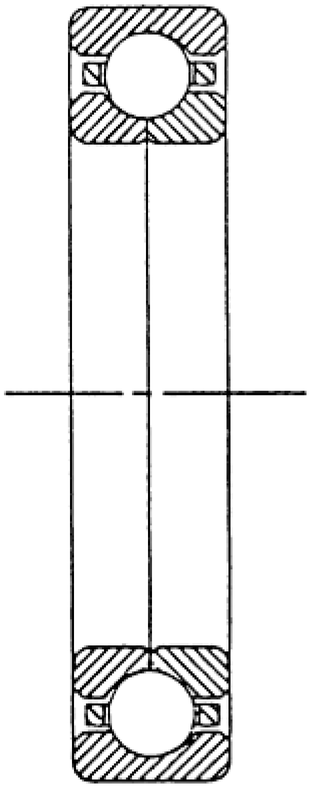


Fig. 7 Typical construction in a split inner-ring ball bearing

Thrust Ball Bearings

Thrust ball bearings (contact angles from 45 to $\leq$90°) are designed to carry predominantly axial load; 90° contact angle thrust bearings (Fig. 8) cannot carry any radial load. Owing to substantial amounts of ball spinning, they are limited to slow- to moderate-speed operation unless means to remove high friction heat associated with high-speed operation is provided.

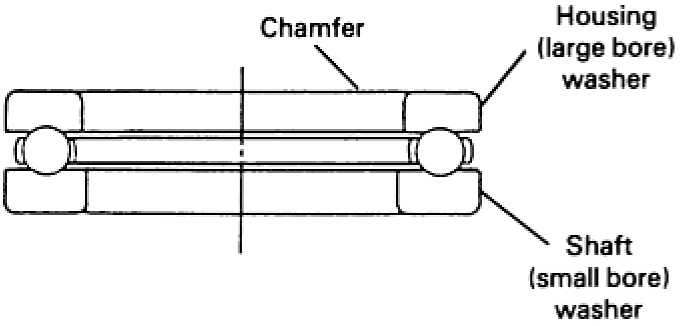


Fig. 8 Thrust ball bearing designed to carry axial load only due to 90° contact angle

Roller Bearings

Radial cylindrical roller bearings, tapered roller bearings, radial spherical roller bearings, and radial needle roller bearings comprise the radial roller bearing group.

Radial Roller Bearings. The basic type of radial roller bearing is the single-row cylindrical roller bearing, which consists of each rolling element having the basic form of a cylinder (Fig. 9). Accurately manufactured bearings have

slightly crowned rollers (Fig. 10) so that, except under extremely heavy loading and/or bearing misalignment, edge loading and its deleterious effects on bearing endurance are avoided. The bearings are capable of supporting very heavy radial loading and operating at very high speeds. They tend to be sensitive to misalignment; only a few minutes of misalignment can be tolerated without a substantial reduction in endurance. In general, the bearings permit axial float of the shaft because neither the inner ring nor the outer ring usually has any guide flanges. When both rings have guide flanges (Fig. 11), however, the bearing can support some thrust load in conjunction with a greater amount of applied radial load. The bearing can be assembled with pressed metal; precision-machined brass, bronze, or steel; or molded plastic cages, depending on the application. These bearings are manufactured in two-row and four-row versions for greater load-carrying capacity.

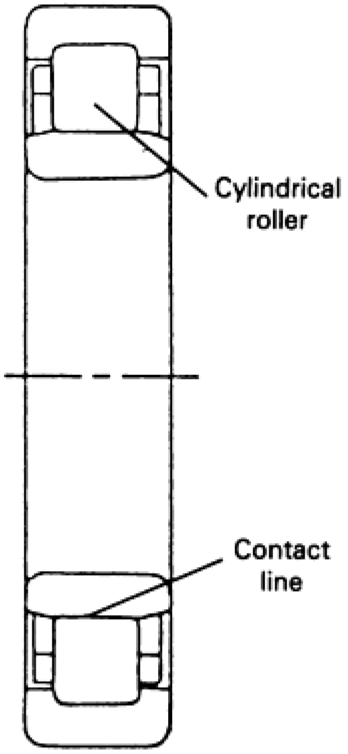


Fig. 9 Radial roller bearing having cylindrical roller elements

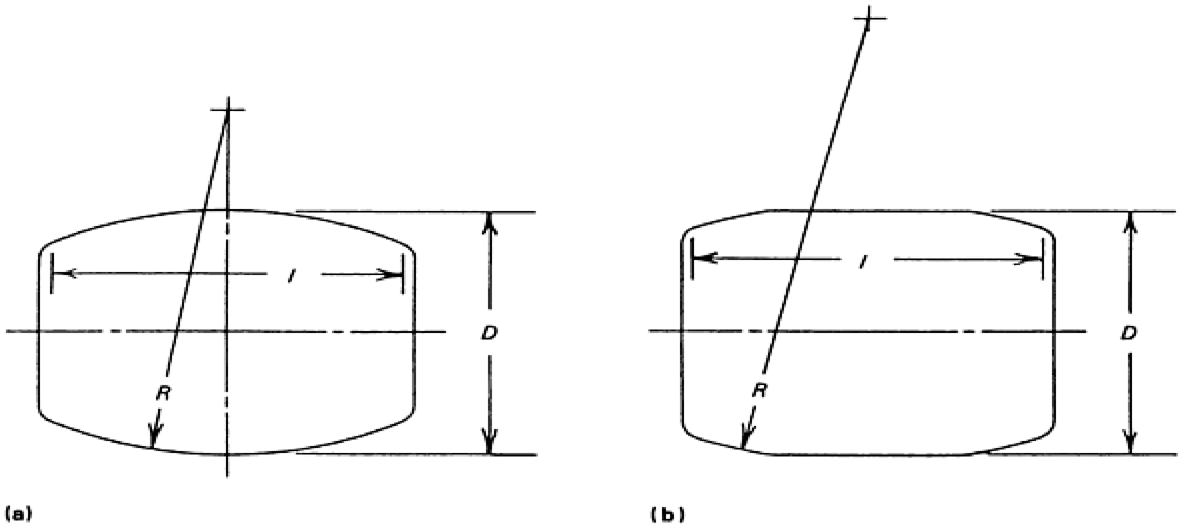


Fig. 10 Crowned roller geometry used in radial roller bearings. (a) Fully crowned (spherical roller). (b) Partially

crowned (cylindrical roller). Roller length, *l;* roller diameter, *D;* crown radius, *R*

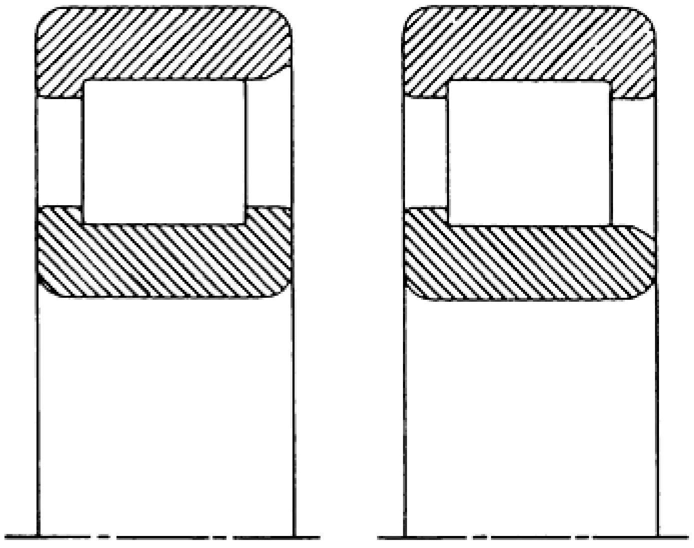


Fig. 11 Radial cylindrical roller bearing with guide flanges on both inner and outer rings

Tapered Roller Bearings. A common form of radial roller bearing is the tapered roller bearing. As shown in Fig. 12, the basic form of the tapered roller is a portion of a cone; generally the rollers are crowned. The outer and inner raceway surfaces are also portions of cones, the outer ring being called the cup and the inner ring the cone. Because the raceways are angled relative to the shaft, radial tapered roller bearings can carry combined radial and axial loading; very heavy loading can be supported. To provide increased load-carrying capacity, the bearings may be obtained in two-row and four-row units.

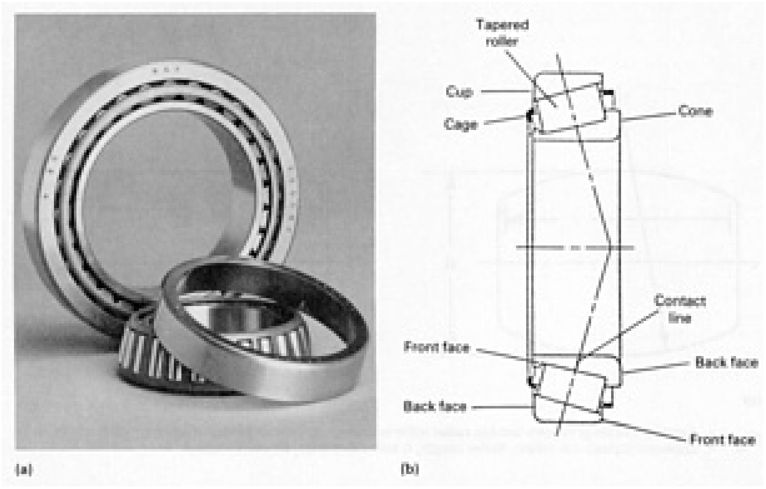


Fig. 12 Key components of a tapered roller bearing. (a) Photograph showing cone assembly and cup of the typical roller bearing shown in background. (b) Schematic of a cross section of a tapered roller bearing

The bearings are sensitive to misalignment, tending to have substantially reduced endurance under only a few minutes of misalignment. Generally, operating misalignment is limited to $\leq$3 min to achieve expected performance. Because the rolling-contact surfaces are portions of cones, the rollers are driven against a fixed guide flange during operation, usually on the inner ring. The contacts between the roller ends and the guide flange, sometimes called the large end rib, are sliding contacts creating substantial friction as compared to the raceway rolling contacts; therefore, the bearings cannot

operate at high speed without a lubrication system capable of removing the friction heat thereby generated. Generally, the bearing cages are manufactured from stamped or pressed steel.

Radial spherical roller bearings are generally two-row bearings having one raceway that is a portion of a sphere; therefore, the bearings are internally self-aligning. In their most common form, the bearings have an outer raceway that is a spherical surface, and the rollers are symmetrical or barrel-shaped (Fig. 13). In this bearing design, a floating flange is usually employed to provide roller guidance; in another design using asymmetrical rollers, the rollers (as in tapered roller bearings) are driven against, and guided by, a fixed axial central flange. Some bearings with symmetrical rollers also use a fixed central guide flange. Depending on the bearing dimension series, from 1 to 2.5° of misalignment can be accommodated without detriment to bearing endurance. Because they have design contact angles that are >0° (usually >10°), double-row bearings can support combinations of very heavy radial and thrust loading. Conversely, a single-row spherical roller bearing has a nominal 0° contact angle and can only support a small amount of thrust load in conjunction with a much greater radial load.

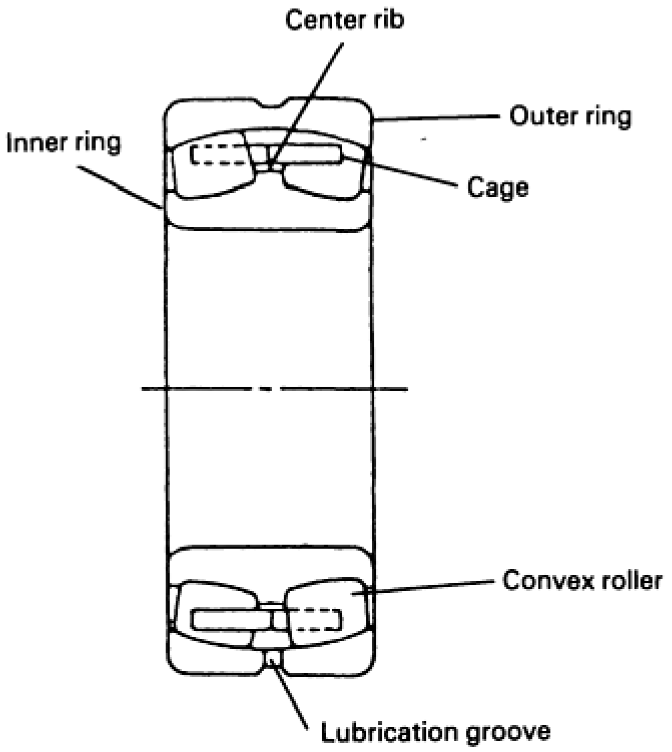


Fig. 13 Typical construction in a radial spherical roller bearing

Because of the relatively close matching of the curved roller profile to the raceways (that is, close osculations to achieve high load-carrying capacity) and bearing contact angle >0°, spherical roller bearings operate with substantial sliding in the roller-raceway contacts. Thus, the bearings are constrained to operate at relatively slow speeds unless lubrication systems are employed that can effectively remove the substantial amounts of friction heat generated. Cages in spherical roller bearings are usually either pressed steel or machined brass or bronze. Double-row bearings may be assembled with a cage for each row or with a one-piece cage.

Radial needle roller bearings (Fig. 14) are radial cylindrical roller bearings whose rollers have length/diameter ratios $\gg$1 (for example, roller length is 3*D* to 10*D*, where *D* is the roller diameter); roller diameter is generally not greater than 5 mm (0.2 in.). The bearings are used in moderate radial load and speed applications where radial space is at a premium. Needle roller bearings may be assembled as full complement bearings (that is, with no cage and a drawn steel cup outer ring or with machined and ground inner and outer rings and a pressed steel or molded plastic cage). Sometimes the bearings are supplied as outer ring-roller cage assemblies with rollers that ride directly on a properly hardened shaft.

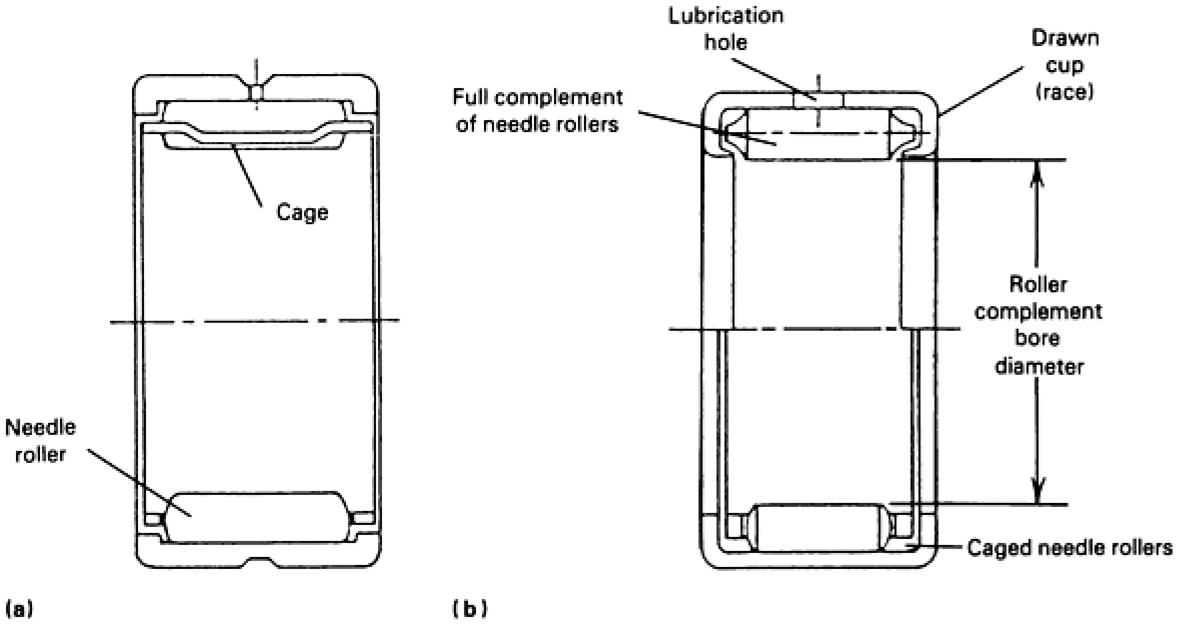


Fig. 14 Two types of radial needle roller bearing. (a) Machined ring with cage. (b) Drawn cup

The allowable operating speed of a full complement assembly is less than that of a bearing having precision-machined rings and cage; the allowable operating speeds of needle roller bearings are less than those of standard cylindrical roller bearings owing to the tendency of longer rollers to skew and cause increased friction heating.

Thrust Roller Bearings

Thrust roller bearings are available in cylindrical roller, needle roller, tapered roller, and spherical roller versions.

Thrust cylindrical roller bearings (Fig. 15) are 90° contact angle bearings designed to support axial load only. The rollers are similar to those of radial cylindrical roller bearings; however, because of the high contact angle, they operate with considerable sliding motion, causing substantial friction heating and limiting use of the bearings to slow-speed applications. To minimize friction, several rollers of relatively small length-diameter ratio may be used in a single-cage pocket as compared to use of a single roller of greater length.

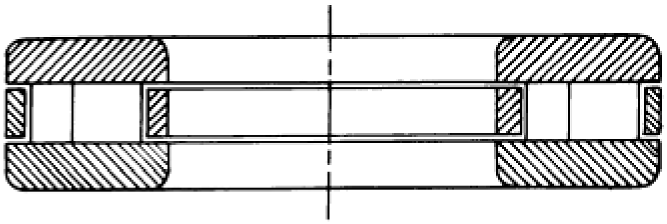


Fig. 15 Thrust cylindrical roller bearing limited to use with axial loads only due to its 90° contact angle

Thrust needle roller bearings operate with characteristics similar to those of thrust cylindrical roller bearings.

Thrust tapered roller bearings have steep contact angles to accommodate primarily axial load. However, in combination with relatively heavier axial load, thrust tapered roller bearings can also support radial load. Similar to radial tapered roller bearings, thrust tapered roller bearings are limited to relatively slow-speed operational unless means to remove friction heating are employed. This type of bearing is similarly sensitive to even small amounts of misalignment.

Spherical roller thrust bearings (Fig. 16) are usually single-row bearings with an outer raceway that is a portion of a sphere and with asymmetrical rollers that are of a continuously curved profile. They can support heavy combined thrust and radial loading, provided that thrust loading predominates and is sufficient to ensure the loading of all rollers. They are internally self-aligning and, depending on specific design, can accept 2 to 3° misalignment without detriment to endurance. They are generally constrained to relatively slow operation because of roller-raceway and roller end-flange friction heating unless means to remove such heat are provided in the application.

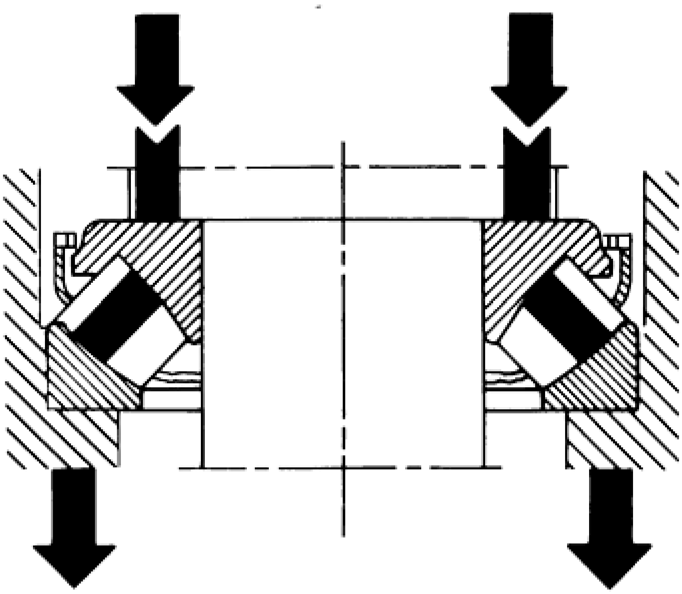


Fig. 16 Spherical roller thrust bearing, which incorporates asymmetrical rollers that tolerate a misalignment of ±3°

Bearing Component Materials

Rolling-Contact Component Steels. Most modern ball and roller bearing rings, balls, and rollers are manufactured from vacuum-processed AISI 52100, a high-carbon through-hardening steel, heat treated to at least 58 HRC. High-quality grades of this steel are required for ball and roller bearings to achieve the standard load rating generally published in the catalogs of bearing manufacturers for each bearing. The heat-treated steel is highly resistant to rolling-contact fatigue, but is considered brittle and susceptible to fracture in heavy load applications where ring bending can occur.

Some bearing types, principally tapered roller bearings, are fabricated from case-hardening steels such as SAE 4320, 4620, and 8620. These low-carbon steels are carburized (that is, heat treated in a carbonaceous atmosphere), such that the case, in which carbon is absorbed (surface and near-surface material), is hardened to $\geq$58 HRC while the remaining core is hardened to $\sim$45 HRC. This provides a case that resists rolling-contact fatigue and a core that resists fracture. The carburized bearing material has a fracture toughness of $\sim$40 MPA $\sqrt{\text{m}}$(36 ksi $\sqrt{\text{in.}}$), approximately twice that of through-hardened 52100 steel. Bearing components fabricated from both 52100 and the case-hardening steels cited above can operate continuously at temperatures up to 160 °C (320 °F) without any reduction of fatigue endurance capability.

For bearings that must operate in corrosive environments (for example, water and salt-laden atmospheres), bearings with rings and rolling elements manufactured from AISI 440C, a stainless steel, are frequently used. Instrument and miniature ball bearings, which must have smooth uninterrupted minimum friction operation at all times, are fabricated from 440C material. This high-carbon chromium-rich steel can be through-hardened to 58 HRC; however, bearing fatigue endurance is somewhat less than that of the same bearing fabricated from AISI 52100. Satisfactory bearing operation at temperatures up to 160 °C (320 °F) can be expected.

Bearings that must operate at elevated temperatures are usually manufactured from tool steels, AISI M50 being most prevalent in use. Bearings with rolling-contact components fabricated from M50 are capable of steady-state operation to

at least 300 °C (572 °F), provided that bearing internal design, cage materials, lubricants, and materials of other components associated with the bearing operation can accommodate the operating temperatures. The steel is routinely through-hardened to ≥58 HRC. M50 steel is supplied as consumable vacuum-melted (CVM) M50 and as vacuum induction melted/vacuum arc remelted (VIMVAR) highly homogeneous, minimum inclusion variants that provide exceptional resistance to rolling-contact fatigue in most applications. M50 Nil, a carburizing variant, provides similar properties in addition to increased fracture toughness.

Rolling-Element Ceramics. As an alternative to steel in machine tool bearing applications, hot isostatically pressed silicon nitride (Si_3N_4) ceramic balls are being used. Bearings with steel rings and ceramic balls are called hybrid ball bearings. Because of the lower density of silicon nitride and its increased rigidity, these high-speed bearings tend to operate with less friction, lower temperatures, and increased stiffness than bearings with steel balls. These bearings also tend to be more expensive.

Cage Materials. The most common materials used for rolling bearing cages are low-carbon steels that can be pressed or stamped (for example, AISI 1010 or 1020). Alternatively, stamped cages are fabricated from brass. Such cages can be either rolling-element riding or land riding. In many applications, molded polyamide 6/6 (nylon) is used in place of metal. The nylon is generally supplied with 25% glass fiber filling for appropriate strength and satisfactory continuous operation at temperatures up to 100 °C (212 °F); they can be used at slightly higher temperatures if operation at these temperatures is intermittent. Bearings with plastic cages tend to operate with slightly less friction and can therefore achieve higher operating speeds without special attention to friction heat removal.

Precision-machined cages are manufactured from steel (for example, AISI 4349, brass, bronze, and variations such as silicon-iron-bronze and alcop (aluminum-copper bronze). Bearings with such machined cages are either inner or outer ringland riding (Fig. 17). Bearings with such cages can generally achieve higher operating speeds than bearings with stamped metal cages. Precision cages are also machined from fabric-reinforced phenolic resin for very high-speed operating bearings (for example, for machine tool applications); however, operating temperatures are limited to 120 °C (248 °F).

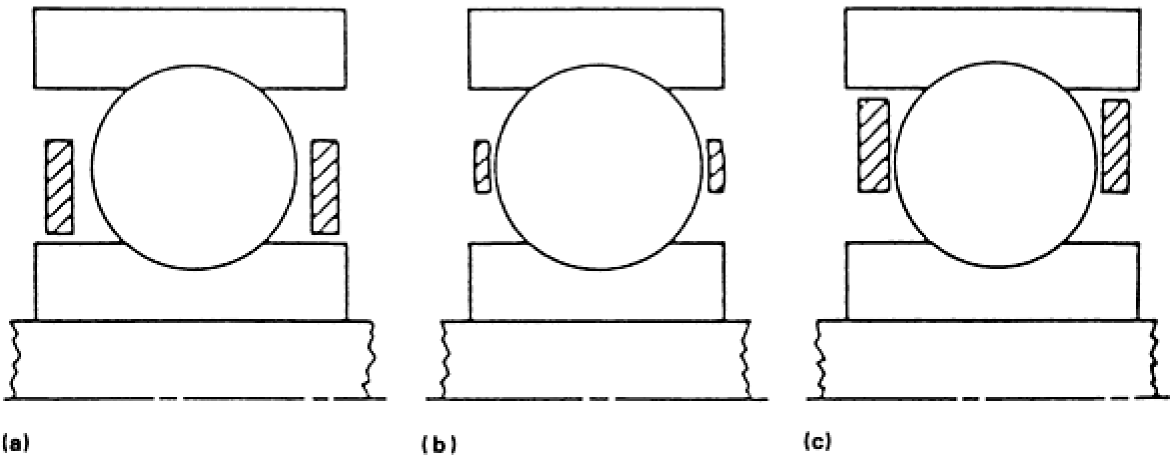


Fig. 17 Schematic showing types of rolling bearing cages. (a) Inner ringland riding. (b) Ball riding. (c) Outer ringland riding

Lubrication Requirements and Methods

To obtain satisfactory performance and rated endurance of rolling bearings, adequate fluid lubrication is necessary. The lubricant is required to:

- Form fluid films between the rolling elements and raceways, rolling elements and cage pockets, cage rails and ring lands, roller ends and abutting flanges, thereby minimizing metal-to-metal contact,

friction, and wear

- Contain chemical additives to minimize rolling-contact surface corrosion and wear
- Transport friction heat away from the bearing

In general, only a small amount of lubricant is required to perform the first two functions. In fact, too much lubricant in the bearing free space or cavity will cause excessive fluid churning, increased bearing friction, and higher operating temperatures. Therefore, the amount of circulating lubricant required to remove friction heat must be carefully determined to avoid overheating and potential early failure.

Grease Lubrication. Most rolling-element applications are lubricated with grease because such lubrication is the most economical and simplest mechanical means to accomplish the task. Greases are combinations of fluid lubricants (oils) with metallic soap thickeners; the oil contained in the grease provides most of its lubricating ability. For grease lubrication to be effective, however, bearing operating speeds must be relatively slow ($<3 \times 10^5\ n \cdot d_m$). This is due to the inability of the grease to dissipate friction heat generation caused by high speed and also its inability to replenish rolling-contact surface fluid lubricant due to insufficient bleeding (flow) rate (lubricant starvation). Heat dissipation from the bearing is achieved mainly by conduction through metal structures.

Many bearings are supplied in the greased-for-life form, in which the grease charge is kept in the bearing by integral seals (Fig. 18). The elastomeric material used as the sealing element must be compatible with the application operating temperatures and environment to obtain satisfactory seal endurance. The amount of grease charge must be carefully metered because excessive grease can cause churning (which produces immediate overheating) and insufficient grease can result in inadequate replenishment of lubricant films (which leads to eventual overheating). When operating temperatures exceed the temperature capability of the grease, aging (oxidation) occurs that reduces the capability of the grease to lubricate. Generally, the charge of grease should occupy from one-third to one-half of the free volume. Lithium soap (thickener) greases are most frequently used; the type of grease used depends on the operating temperatures and application requirements. The upper temperature limit for a grease is determined by the type of thickener. A rule of thumb is that grease life is halved with each 11 °C (20 °F) increase in bearing temperature.

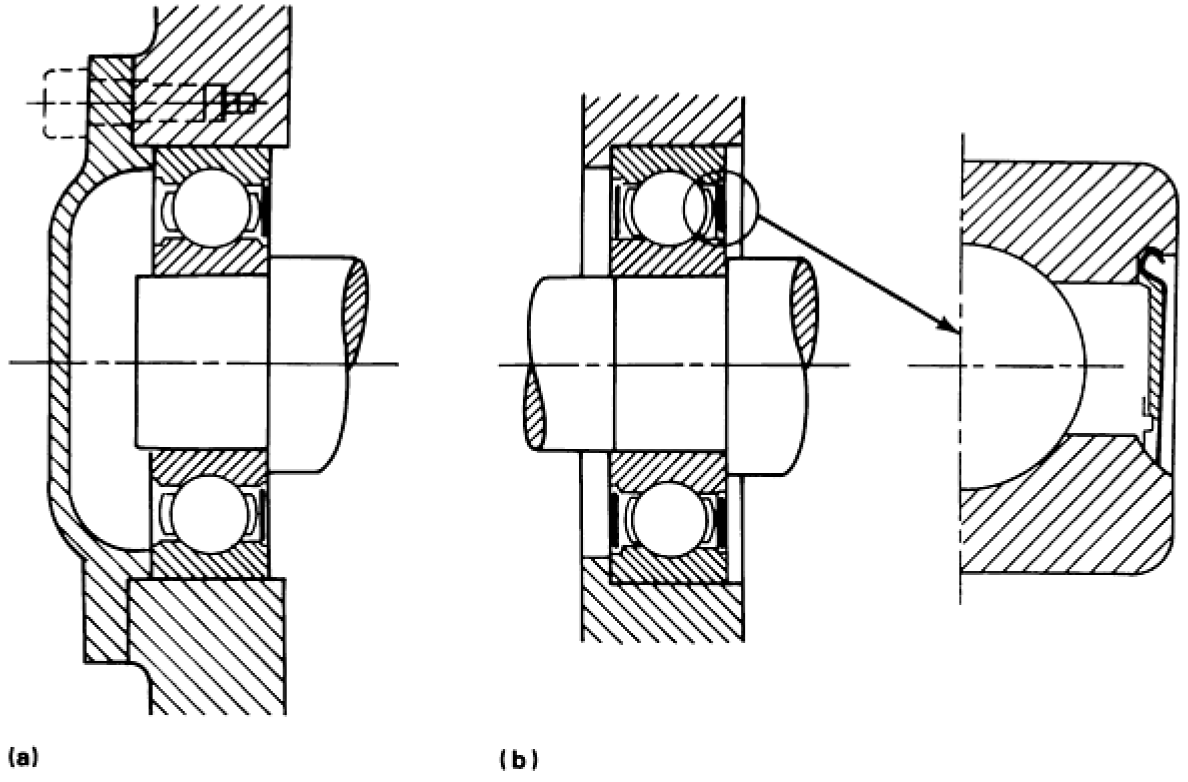


Fig. 18 Location of elastomeric integral seals in a ball bearing. (a) Single-face seal. (b) Dual-face seal

Rolling bearings are frequently mounted in pillow blocks (housings), which have fittings for regreasing (Fig. 19). In this case, the bearings are open and the pillow blocks are supplied with seals to prevent the loss of grease. Regreasing intervals, as recommended by bearing manufacturers, depend on the bearing design and application conditions. When a bearing is regreased, it is essential that the new grease pushes through the bearing and the used grease is effectively purged.

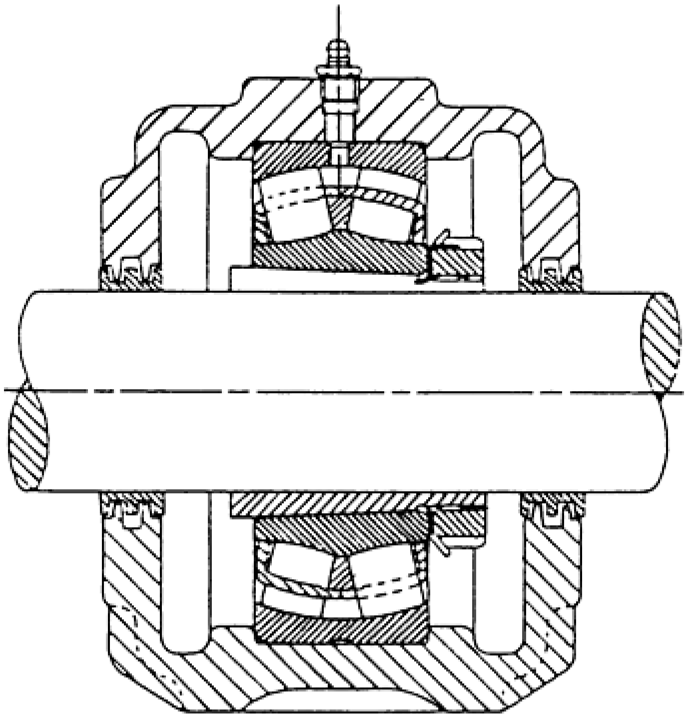


Fig. 19 Location of regreasing fitting in a pillow block mounting

For additional information about the composition and application of grease, see the article "Grease" in this Volume.

Bath Lubrication. Many rolling bearings are effectively operated using oil bath lubrication. For horizontally mounted bearings, a portion of the bearing ball or roller complement is covered by the lubricant bath (Fig. 20). As the cage-rolling element set rotates, each rolling element in succession dips into the bath, carrying sufficient oil with it to adequately lubricate the entire bearing. Lubricant returns by gravity to the bath. The level of the bath must be such that the process is achieved without lubricant starvation at the rolling contacts most remote from the bath. If the lubricant level is too high, excessive lubricant churning occurs, with potential bearing overheating and failure. In general, the lubricant level should not be higher than the center of the lowest rolling element under nonoperating conditions.

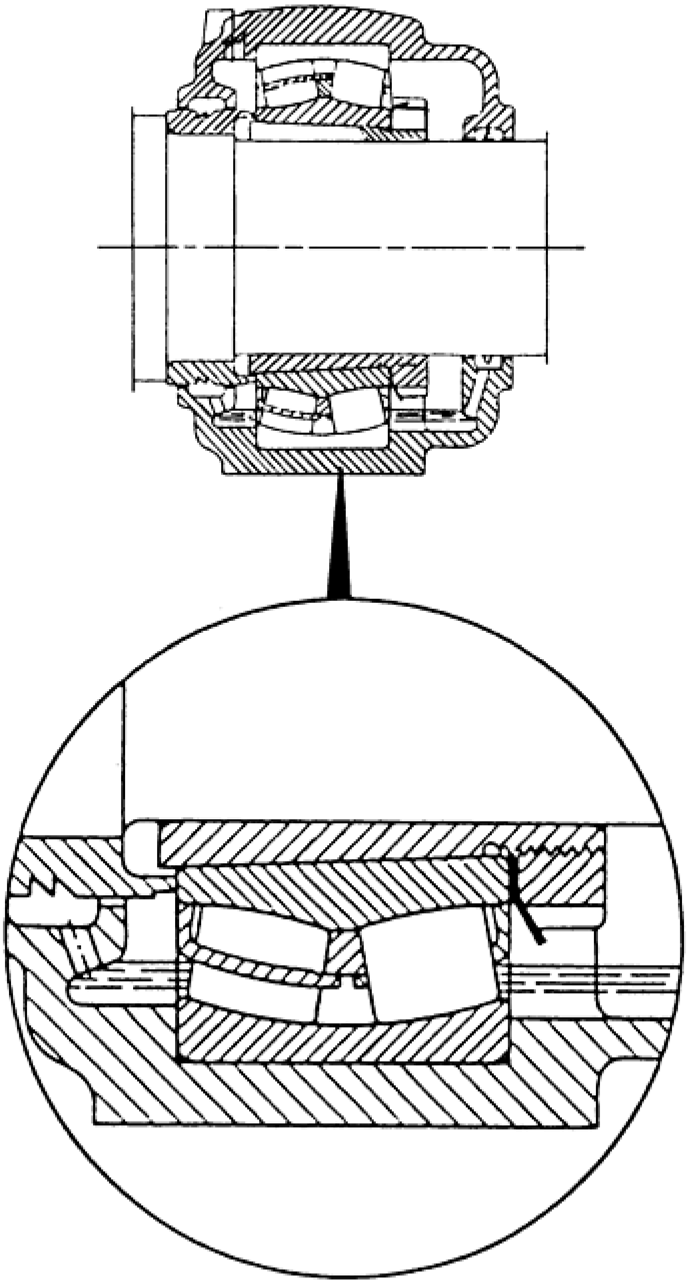


Fig. 20 Oil bath lubrication suitable for low speeds of rotation. When the bearing is not rotating, oil should come to level just below center of lowest ball or roller. When the bearing rotates, oil is drawn up by bearing moving components, transported through the bearing, and returned by gravity to the bath.

For vertically mounted bearing applications, the optimum lubricant levels must be established according to the operating speeds and bearing internal design.

Bath-lubricated bearings can operate at higher speeds than grease-lubricated bearings because the bearing components transfer most of their friction heat to the lubricant by convection. The lubricant subsequently splashes on the bearing housing internal surfaces, transferring heat by mass transfer and convection. This heat is then conducted through the housing structure for dissipation by free or forced convection to the environment.

Circulating-Oil Lubrication. When bearing operating speeds are high and applied loading is very heavy, it is frequently necessary to utilize circulating-oil lubrication. In this case, lubricant that is supplied to the bearing is subsequently passed through a heat exchanger to remove friction heat (that is, reduce the lubricant temperature prior to returning the lubricant to the bearing). In applications where the generation of friction heat is exceptionally great, a jet

supply of lubricant can be used in which one or more jets of lubricant impinges directly on the bearing rotating components (Fig. 21). Where jet supply of lubricant is used, significant amounts of lubricant churning can be expected, with attendant high friction heating. This extra heat must also be removed, increasing the bearing cooling requirements.

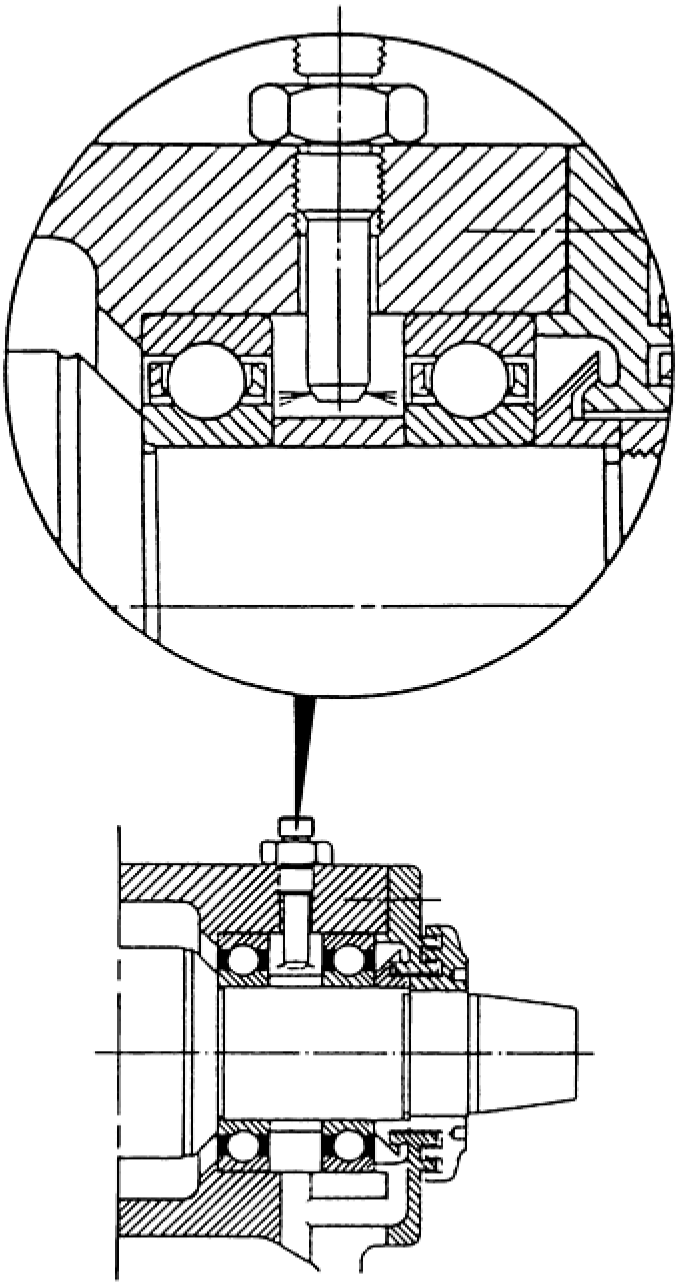


Fig. 21 Oil jet lubrication suitable for high speeds of rotation. Oil is injected into the bearing under pressure to gain access to the bearing interior to remove heat. Oil jet velocity must be high enough to piece air vortex caused by bearing rotation.

Load Ratings

In general, to satisfy application conditions, ball and roller bearings are selected from the catalogs of bearing manufacturers according to basic dynamic and basic static load ratings. For most manufacturers, these ratings are based on internationally established standard methods of calculation (Ref 2, 3, 4, 5). Basic dynamic load rating in conjunction with the applied loading is used to estimate the ability of the bearing to withstand the surface fatigue of the components in rolling contact. Basic static load rating in conjunction with the statically applied load is used to determine the ability of

the bearing to resist permanent indentations. Frequently, many other criteria must be considered to establish the suitability of a given rolling bearing for a given application (for example, elastic deflection, friction torque, friction heat generation, operating temperatures, corrosion resistance, operating vibration, noise level, and son on). Nevertheless, most bearings are selected for given applications based on their resistance to rolling-contact fatigue and permanent deformation, and based on the load ratings published in the catalogs of manufacturers.

Fatigue failure of the surfaces in rolling contact was selected as the endurance criterion for ball and roller bearings because it was considered that all other modes of bearing failure could be prevented if the bearing rolling contact components were:

- Accurately manufactured from high-quality bearing steel
- Properly lubricated
- Keep free of contaminants
- Properly mounted and operated in the application

Bearing endurance as limited by rolling-contact fatigue represented the longest period of time a bearing could reliably perform its intended function. The failure mode manifested itself as a spall or flaking of material from a rolling-contact surface. This failure mode is considered a form of wear because material is removed from the rolling-contact surfaces (Fig. 22). Moreover, bearing endurance is a statistical phenomenon; that is, apparently identical roller bearings subjected to apparently identical conditions of operation will not achieve identical failure endurance. Instead, they will tend to fail according to a statistical distribution (Fig. 23). A Weibull distribution (Ref 6) has been shown to adequately approximate the bearing fatigue life dispersion. For ball bearings, it has been demonstrated that the median life, L_{50}, is approximately equal to five times the rating life, L_{10}:

$$L_{50} = 5L_{10} \qquad \textbf{(Eq 1)}$$

Standard ball and roller bearing dynamic load ratings are based on the work of Lundberg and Palmgren (Ref 7, 8). Their method of determining basic dynamic load ratings was based on rolling-contact fatigue testing of ball and roller bearings during the 1930s and 1940s. Because of the substantial improvements in bearing steels, bearing steel processing, and bearing manufacturing methods since that time, the standard calculation methods have been modified several times to accommodate the improvements. The basic static load rating is based on the work of Lundberg *et al.* (Ref 9); the current standard method reflects relatively recent improvement in the accuracy of determining permanent deformations during testing. Both basic dynamic load ratings and basic static load ratings have undergone substantial increases since the original rating methods were established. Because bearing load ratings have increased substantially over time, it follows that to resist specific loading, smaller bearings are now being employed (that is, downsizing of bearings has occured).

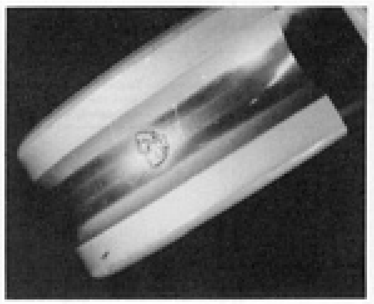


Fig. 22 Fatigue spall centered on a ball bearing race-way

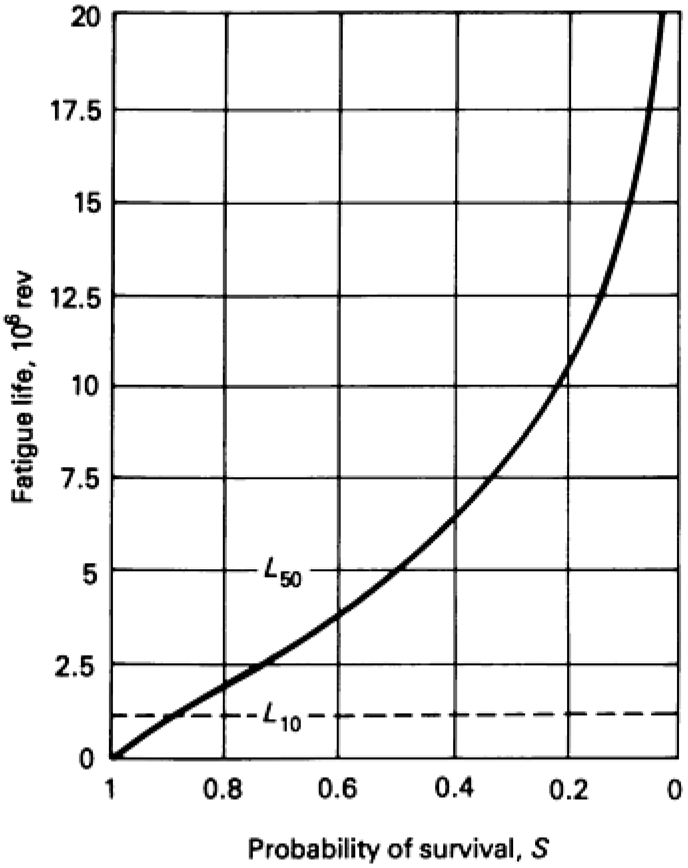


Fig. 23 Plot of fatigue life versus probability of survival to indicate rolling bearing fatigue life distribution.

Basic Load Rating

Basic load rating (also called basic dynamic load rating or basic dynamic capacity) is defined as that load which 90% of a group of apparently identical rolling bearings will survive (rolling-contact fatigue) with a life of one million revolutions. For radial ball bearings, the International Standards Organization (ISO) (Ref 2) gives the following equations:

$$C_r = b_m f_c (i \cdot \cos \alpha)^{0.7} Z^{2/3} D^{1.8} \quad D \leq 25.4 \text{ mm} \qquad \textbf{(Eq 2)}$$

$$C_r = 3.647 b_m f_c (i \cdot \cos \alpha)^{0.7} Z2/3D^{1.4} \quad D > 25.4 \text{ mm} \qquad \textbf{(Eq 3)}$$

in which

C_r = basic radial load rating (in newtons)

b_m = rating factor for contemporary, normally used material and manufacturing quality, the value of which varies with the bearing type and design

f_c = factor that depends on the composition of the material, the geometry of the bearing components, and the accuracy to which the various components are made

- i = number of rows of balls
- α= nominal contact angle (in degrees)
- Z = number of balls per row
- D = ball diameter (in mm)

b_m can be obtained from Table 1 for different bearing types, and f_c can be obtained from Table 2 as a function of the geometry parameter ($D \cdot \cos \alpha$)/d, in which d is the bearing pitch diameter (in mm). The tabular data pertain strictly to radial and angular-contact bearings having inner ring groove radii $\leq$0.52D and outer ring groove radii $\leq$0.53D and to self-aligning bearings having inner ring groove radii $\leq$0.53D.

Table 1 b_m factors for selected rolling bearing types

Bearing type	b_m
Ball bearings	
Radial and angular-contact ball	1.3
Filling slot ball	1.1
Self-aligning ball	1.3
Thrust ball	1.3
Roller bearings	
Cylindrical roller	1.1
Tapered roller	1.1
Needle roller with machined rings	1.1
Drawn cup needle roller	1
Spherical roller	1.15
Cylindrical roller thrust	1
Needle roller thrust	1
Tapered roller thrust	1
Spherical roller thrust	1.15

Table 2 f_c factors for radial ball bearings and angular-contact ball bearings

$(D \cdot \cos \alpha)/d$ (a)	f_c			
	Single-row radial contact groove ball bearings and single-row and double-row angular contact groove ball bearings	**Double-row radial contact groove ball bearings**	**Single-row and double-row self-aligning ball bearings**	**Single-row radial contact separable ball bearings (magneto bearings)**
0.01	29.1	27.5	9.9	9.4
0.02	35.8	33.9	12.4	11.7
0.03	40.3	38.2	14.3	13.4
0.04	43.8	41.5	15.9	14.9
0.05	46.7	44.2	17.3	16.2
0.06	49.1	46.5	18.6	17.4
0.07	51.1	48.4	19.9	18.5
0.08	52.8	50	21.1	19.5
0.09	54.3	51.4	22.3	20.6
0.10	55.5	52.6	23.4	21.5
0.11	56.6	53.6	24.5	22.5
0.12	57.5	54.5	25.6	23.4
0.13	58.2	55.2	26.6	24.4
0.14	58.8	55.7	27.7	25.3
0.15	59.3	56.1	28.7	26.2
0.16	59.6	56.5	29.7	27.1
0.17	59.8	56.7	30.7	27.9
0.18	59.9	56.8	31.7	28.8
0.19	60	56.8	32.6	29.7
0.20	59.9	56.8	33.5	30.5
0.21	59.8	56.6	34.4	31.3
0.22	59.6	55.5	35.2	32.1
0.23	59.3	56.2	36.1	32.9
0.24	59	55.9	36.8	33.7
0.25	58.6	55.5	37.5	34.5
0.26	58.2	55.1	38.2	35.2
0.27	57.7	54.6	38.8	35.9
0.28	57.1	54.1	39.4	36.6
0.29	56.6	53.6	39.9	37.2
0.30	56	53	40.3	37.8
0.31	55.3	52.4	40.6	38.4
0.32	54.6	51.8	40.9	38.9
0.33	53.9	51.9	41.1	39.4
0.34	53.2	50.4	41.2	39.8
0.35	52.4	49.7	41.3	40.4
0.36	51.7	48.9	41.3	40.4
0.37	50.9	48.2	41.2	40.7
0.38	50	47.4	41	40.8
0.39	49.2	46.6	40.7	40.9
0.40	48.4	45.8	40.4	40.9

(a) f_c values for intermediate values of $(D \cdot \cos \alpha)/d$ are obtained by linear interpolation.

Standard Bearing Internal Geometry

The ratio of the raceway curvature radius (ring groove radius) to the ball diameter, $f = r/D$, together with the parameter $(D \cdot \cos \alpha)/d$, internal radial clearance, and applied load determine the normal stresses in the rolling element-raceway contacts. The standard formulas for basic radial load rating are based on the nominal clearance that produces a radial load distribution among the rolling elements (Fig. 24). Normal stresses depend on the contact geometry and rolling-element load; hence, maximum stress occurs colinear with the applied radial load. Moreover, the contact areas are elliptical in shape, and the distribution of stress in each contact is as shown in Fig. 25. In most bearing applications, maximum normal stress in the contacts rarely exceeds 2000 MPa (290 ksi); however, maximum stresses as high as 3300 MPa (475 ksi) can be carried (by properly hardened steel bearings) without significant permanent deformation of the contact geometry. If bearing raceway groove curvatures are different from those indicated above, then the standard load rating formulas cannot be used. If the raceway curvature radius is greater than standard, normal stresses will be greater than standard and the basic load rating can be considerably less than standard. Also, if a bearing has an internal radial clearance greater than

standard, maximum rolling-element load will be greater than standard, and the basic load rating will accordingly be less than standard. Harris (Ref 10) provides methods for rating bearings with nonstandard internal geometries.

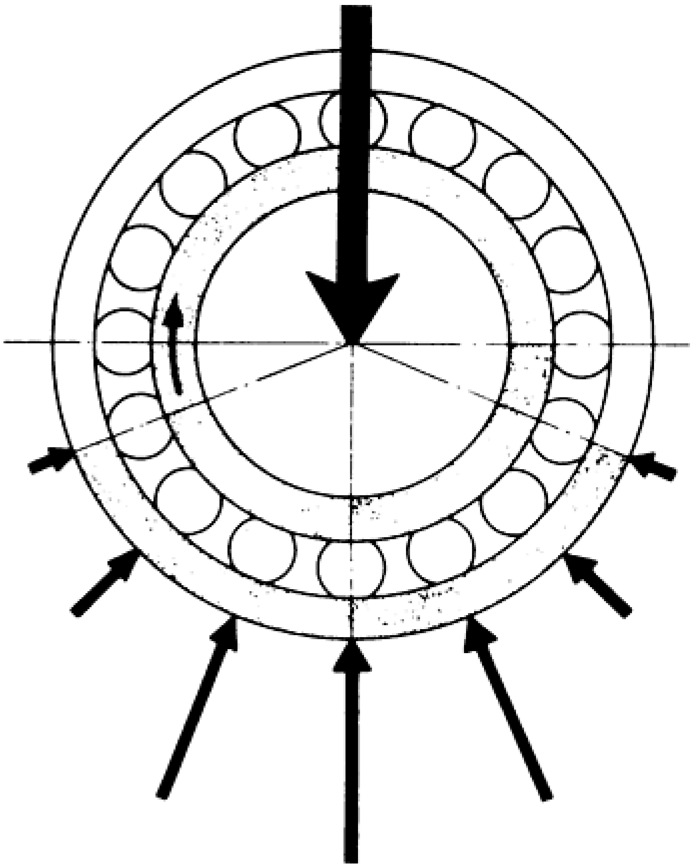


Fig. 24 Distribution of radial load among rolling elements in a bearing

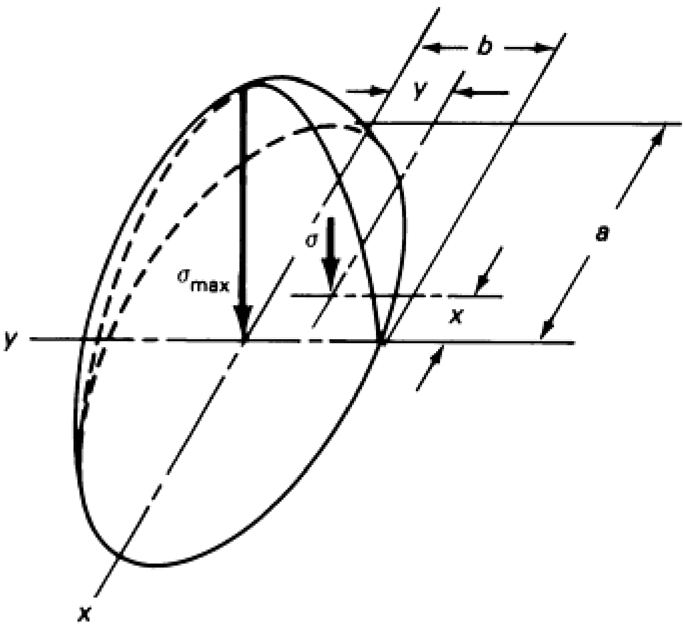


Fig. 25 Normal (Hertzian) contact stress distribution obtained over an elliptical rolling element-raceway contact area in a bearing. σ_{max} is the maximum normal contact stress at the center of contact and σ is the contact

stress anywhere else.

For thrust ball bearings the following equations pertain:

$$C_a = b_m f_c Z^{2/3} D^{1.8} \quad D \leq 25.4 \text{ mm}; \alpha = 90° \qquad \textbf{(Eq 4)}$$

$$C_a = b_m f_c (\cos \alpha)^{0.7} (\tan \alpha) Z^{2/3} D^{1.8} \quad D \leq 25.4 \text{ mm}; \alpha \neq 90° \qquad \textbf{(Eq 5)}$$

$$C_a = 3.647 \; b_m f_c Z^{2/3} D^{1.4} \quad D > 25.4 \text{ mm}; \alpha = 90° \qquad \textbf{(Eq 6)}$$

$$C_a = 3.647 b_m f_c (\cos \alpha)^{0.7} (\tan \alpha) Z^{2/3} D^{1.4} \quad D > 25.4 \text{ mm}; \alpha \neq 90° \qquad \textbf{(Eq 7)}$$

where Z is the number of balls carrying load in one direction. As mentioned earlier in the section "Basic Load Rating" in this article, b_m can be obtained from Table 1; values of f_c are given in Table 3. The tabular data pertain strictly to bearings with ring groove radii $\leq 0.54D$. For bearings with two or more rows:

$$C_a = \sum_{1}^{i} Z_n \cdot \left[\sum_{1}^{i} \left(\frac{Z_n}{C_{an}} \right)^{10/3} \right]^{-3/10} \qquad \textbf{(Eq 8)}$$

Table 3 f_c factors for thrust ball bearings

D/d[(a)]	f_c, α = 90°	$(D \cdot \cos \alpha)/d$[(a)]	f_c		
			α = 45°[(b)]	**α = 60°**	**α = 75°**
0.01	36.7	0.01	42.1	39.2	37.3
0.02	45.2	0.02	51.7	48.1	45.9
0.03	51.1	0.03	58.2	54.2	51.7
0.04	55.7	0.04	63.3	58.9	56.1
0.05	59.5	0.05	67.3	62.6	59.7
0.06	62.9	0.06	70.7	65.8	62.7
0.07	65.8	0.07	73.5	68.4	65.2
0.08	68.5	0.08	75.9	70.7	67.3
0.09	71	0.09	78	72.6	69.2
0.10	73.3	0.1	79.7	74.2	70.7
0.11	75.4	0.11	81.1	75.5	. . .
0.12	77.4	0.12	82.3	76.6	. . .
0.13	79.3	0.13	83.3	77.5	. . .
0.14	81.1	0.14	84.1	78.3	. . .
0.15	82.7	0.15	84.7	78.8	. . .
0.16	84.4	0.16	85.1	79.2	. . .
0.17	85.9	0.17	85.4	79.5	. . .
0.18	87.4	0.18	85.5	79.6	. . .
0.19	88.8	0.19	85.5	79.6	. . .
0.20	90.2	0.2	85.4	79.5	. . .
0.21	91.5	0.21	85.2	. . .	. . .
0.22	92.8	0.22	84.9	. . .	. . .
0.23	94.1	0.23	84.5	. . .	. . .
0.24	95.3	0.24	84	. . .	. . .
0.25	96.4	0.25	83.4	. . .	. . .
0.26	97.6	0.26	82.8	. . .	. . .
0.27	98.7	0.27	82	. . .	. . .
0.28	99.8	0.28	81.3	. . .	. . .
0.29	100.8	0.29	80.4	. . .	. . .

0.30	101.9	0.3	79.6	...	...
0.31	102.9	...	...	...	...
0.32	103.9	...	...	...	...
0.33	104.8	...	...	...	...
0.34	105.8	...	...	...	...
0.35	106.7	...	...	...	...

(a) f_c values for D/d or $(D \cdot \cos \alpha)/d$ and α other than shown in the table are obtained by linear interpolation or extrapolation.

(b) Values for $\alpha = 45°$ permit interpolation of values for $45° < \alpha < 60°$.

Standard Bearing Contact Geometry

As with radial ball bearings, standard formulas for thrust ball bearing load ratings apply only for bearings having standard internal geometry. Moreover, it is presumed that the thrust load carried by each ball is identical (that is, the internal load distribution shown in Fig. 26 applies).

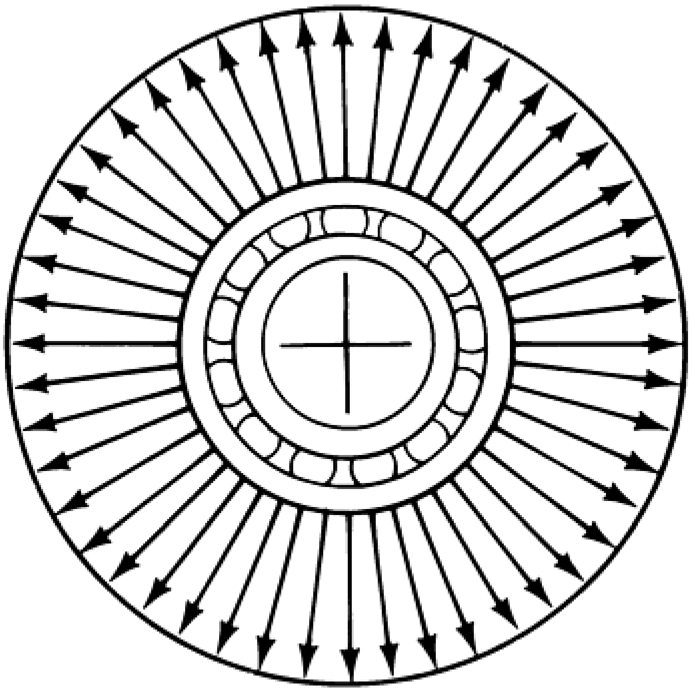


Fig. 26 Distribution of (simple) thrust load among rolling elements in a bearing

For radial roller bearings:

$$C_r = b_m f_c (il \cdot \cos \alpha)^{7/9} Z^{3/4} D^{29/27} \qquad \textbf{(Eq 9)}$$

where l is the roller effective length (in mm). Roller effective length is the total length of the roller minus corner radii or raceway undercuts, whichever length is smaller. Values of b_m are given in Table 1; f_c values are given in Table 4.

Table 4 Maximum f_c factors for radial roller bearings

$(D \cdot \cos \alpha)/d$[a]	f_c
0.01	52.1
0.02	60.8
0.03	66.5
0.04	70.7
0.05	74.1
0.06	76.9
0.07	79.2
0.08	81.2
0.09	82.8
0.10	84.2
0.11	85.4
0.12	86.4
0.13	87.1
0.14	87.7
0.15	88.2
0.16	88.5
0.17	88.7
0.18	88.8
0.19	88.8
0.20	88.7
0.21	88.5
0.22	88.2
0.23	87.9
0.24	87.5
0.25	87
0.26	86.4
0.27	85.8
0.28	85.2
0.29	84.5
0.30	83.8

(a) f_c values for intermediate values of $(D \cdot \cos \alpha)/d$ are obtained by linear interpolation.

Similarly, for thrust roller bearings:

$$C_a = b_m f_c l^{7/9} Z^{3/4} D^{29/27} \quad \alpha = 90° \qquad \textbf{(Eq 10)}$$

$$C_a = b_m f_c (l \cdot \cos \alpha)^{7/9} (\tan \alpha) Z^{3/4} D^{29/27} \quad \alpha \neq 90° \qquad \textbf{(Eq 11)}$$

Values of b_m are given in Table 1; f_c values are given in Table 5. For bearings with two or more rows of rollers:

$$C_a = \sum_1^i (Z_n l_n) \cdot \left[\sum_1^i \left(\frac{Z_n l_n}{C_{an}} \right)^{9/2} \right]^{-2/9} \qquad \textbf{(Eq 12)}$$

Table 5 Maximum f_c factors for thrust roller bearings

D/d [a]	f_c, α = 90°	$(D \cdot \cos \alpha)/d$ [a]	f_c		
			α = 50° [b]	**α = 65°** [c]	**α = 80°** [d]
0.01	105.4	0.01	109.7	107.1	105.6
0.02	122.9	0.02	127.8	124.7	123
0.03	134.5	0.03	139.5	136.2	134.3
0.04	143.4	0.04	148.3	144.7	142.8
0.05	150.7	0.05	155.2	151.5	149.4
0.06	156.9	0.06	160.9	157	154.9
0.07	162.4	0.07	165.6	161.6	159.4
0.08	167.2	0.08	169.5	165.5	163.2
0.09	171.7	0.09	172.8	168.7	166.4
0.10	175.7	0.10	175.5	171.4	169
0.11	179.5	0.11	177.8	173.6	171.2
0.12	183	0.12	179.7	175.4	173
0.13	186.3	0.13	181.1	176.8	174.4
0.14	189.4	0.14	182.3	177.9	175.5
0.15	192.3	0.15	183.1	178.8	176.3
0.16	195.1	0.16	183.7	179.3	. . .
0.17	197.7	0.17	184	179.6	. . .
0.18	200.3	0.18	184.1	179.7	. . .
0.19	202.7	0.19	184	179.6	. . .
0.20	205	0.20	183.7	179.3	. . .
0.21	207.2	0.21	183.2	. . .	. . .
0.22	209.4	0.22	182.6	. . .	. . .
0.23	211.5	0.23	181.8	. . .	. . .
0.24	213.5	0.24	180.9	. . .	. . .
0.25	215.4	0.25	179.8	. . .	. . .
0.26	217.3	0.26	178.7	. . .	. . .
0.27	219.1	. . .	. . .	. . .	. . .
0.28	220.9	. . .	. . .	. . .	. . .
0.29	222.7	. . .	. . .	. . .	. . .
0.30	224.3	. . .	. . .	. . .	. . .

(a) f_c values for intermediate values of D/d or $(D \cdot \cos \alpha)/d$ are obtained by linear interpolation.

(b) Applicable for 45° < α < 60°.

(c) Applicable for 60° $\leq \alpha$ < 75°.

(d) Applicable for 75° $\leq \alpha$ < 90°

Standard Roller Geometry

Use of Eq 9, 10, 11, and 12 presumes that the rollers are crowned as in Fig. 10 so that under loads as heavy as 4*C*, where *C* is the bearing capacity, stress concentrations due to edge loading (Fig. 27) in the contacts are avoided. If appropriate crowning is not employed, basic load ratings can be significantly reduced from standard. With regard to nonstandard bearing internal clearance, the load rating situation is identical to that for ball bearings (that is, excessive clearance will result in reduced load-carrying capacity).

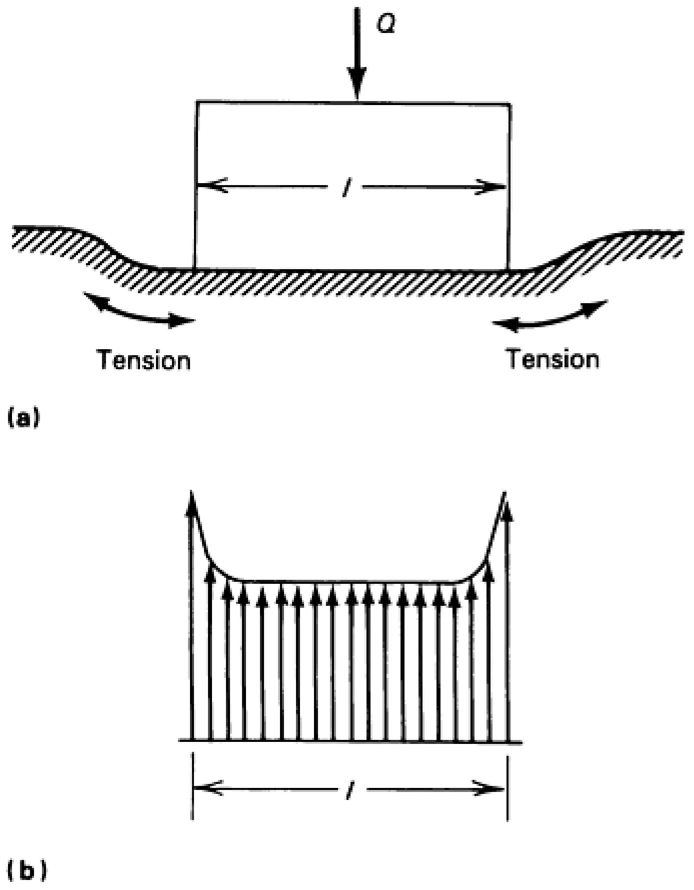


Fig. 27 Roller axial load distribution associated with edge loading. (a) Schematic showing load, *Q*, applied to roller of length, *l*. (b) Load distribution across roller

Bearing Fatigue Life

Using the basic load rating obtained either from a catalog supplied by a bearing manufacturer or from the appropriate equation, bearing rolling-contact fatigue life in a given application can be determined using the following equation:

$$L_{na} = a_1 a_2 a_3 \left[\frac{C}{P}\right]^x \qquad \textbf{(Eq 13)}$$

where

- L_{na} = bearing life (in millions of revolutions)
- a_1 = life adjustment factor for reliability
- a_2 = life adjustment factor for special bearing properties
- a_3 = life adjustment factor for operating conditions
- P = dynamic equivalent load (in N)

In Eq 13, the exponent x = 3 for ball bearings and x = 10/3 for roller bearings; a_1 values for desired levels of reliability can be obtained from Table 6.

Table 6 Life adjustment factor for reliability, a_1

Reliability, %	L_{na}	a_1
90	L_{10a}	1
95	L_{5a}	0.62
96	L_{4a}	0.53
97	L_{3a}	0.44
98	L_{2a}	0.33
99	L_{1a}	0.21

For bearings specified in the catalogs of manufacturers, a_2 factors of 1 are appropriate in the calculation of fatigue life; however, a_2 values >1 may be applicable if the bearings are fabricated from steel of particularly low impurity content or of special analysis. Conversely, if hardness reduction occurs due to a special heat treatment, $a_2 < 1$ should be used. Moreover, special internal design involving increased or reduced uniformity of stress in the rolling contacts can be accommodated by appropriate selection of a_2.

The a_3 operating conditions factor depends strongly on the effectiveness of lubrication and the presence of foreign matter (contaminants) in the application. In the calculation of bearing fatigue life, lubrication is considered normal ($a_3 = 1$) when the lubricant films generated in the rolling element-raceway contacts have thicknesses equal to, or slightly greater than, the composite root-mean-square (rms) roughness of the contacting surfaces. $a_3 < 1$ should be considered when kinematic viscosity of the lubricant at the bearing operating temperature is <13 cSt for ball bearings or <20 cSt for roller bearings and bearing operating speed in rev/min, n, times pitch diameter in mm, d_m, is $<10^4$ ($nd_m < 10^4$). When lubricant films are insufficient to completely separate the rolling-contact surfaces, it is possible that a chemical interaction will occur between the bearing material and the lubricant that will necessitate a reduction in a_2. Accordingly, it cannot be assumed that a reduced value of a_3 can be compensated by use of a steel nominally having $a_2 > 1$. a_3 values > 1 can be considered only where lubrication conditions are so favorable that the probability of failure by surface-oriented distress is virtually eliminated. Bearing manufacturers can recommend appropriate values of a_3.

Equivalent Load. The dynamic equivalent load for a radial bearing is that purely radial load which, if applied to the bearing, would result in the same fatigue life as the actual applied load. For bearings subjected to combined radial and thrust loading, it can be determined from Eq 14:

$$P = XF_r + YF_a \qquad \textbf{(Eq 14)}$$

where F_r is the applied load (in N), and F_a is the applied thrust load (in N). For radial ball and roller bearings, the factors X and Y can be obtained from Tables 7 and 8, respectively.

Table 7 *X* and *Y* factors for radial ball bearings

Bearing type	Relative axial load[(a)(b)]		Single-row bearings				Double-row bearings				*e*
			$F_a/F_r \leq e$		$F_a/F_r > e$		$F_a/F_r \leq e$		$F_a/F_r > e$		
	$(f_oF_a)/C_{or}$	$F_a/(iZD_w^2)$	*X*	*Y*	*X*	*Y*	*X*	*Y*	*X*	*Y*	
Radial contact ball bearings	0.172	0.172	1	0	0.56	2.3	1	0	0.56	2.3	0.19
	0.345	0.345	1	0	0.56	1.99	1	0	0.56	1.99	0.22
	0.689	0.689	1	0	0.56	1.71	1	0	0.56	1.71	0.26
	1.03	1.03	1	0	0.56	1.55	1	0	0.56	1.55	0.28
	1.38	1.38	1	0	0.56	1.45	1	0	0.56	1.45	0.3
	2.07	2.07	1	0	0.56	1.31	1	0	0.56	1.31	0.34
	3.45	3.45	1	0	0.56	1.15	1	0	0.56	1.15	0.38
	5.17	5.17	1	0	0.56	1.04	1	0	0.56	1.04	0.42
	6.89	6.89	1	0	0.56	1	1	0	0.56	1	0.44
	$(f_oiF_a)/C_{or}$	$F_a/(ZD_w^2)$	*X*	*Y*	*X*	*Y*	*X*	*Y*	*X*	*Y*	
Angular-contact groove ball bearings α= 5°	0.173	0.172	1	0	For this type use the X_o, Y_o, and *e* values applicable to single row-radial contact groove ball bearings.		1	2.78	0.78	3.74	0.23
	0.346	0.345	1	0			1	2.4	0.78	3.23	0.26
	0.692	0.689	1	0			1	2.07	0.78	2.78	0.3
	1.04	1.03	1	0			1	1.87	0.78	2.52	. . .
	1.38	1.38	1	0			1	1.75	0.78	2.36	0.36
	2.08	2.07	1	0			1	1.58	0.78	2.13	0.4
	3.46	3.45	1	0			1	1.39	0.78	1.87	0.45
	5.19	5.17	1	0			1	1.26	0.78	1.69	0.5
	6.92	6.89	1	0			1	1.21	0.78	1.63	0.52
α= 10°	0.175	0.172	1	0	0.46	1.88	1	2.18	0.75	3.06	0.29
	0.35	0.345	1	0	0.46	1.71	1	1.98	0.75	2.78	0.32
	0.7	0.689	1	0	0.46	1.52	1	1.76	0.75	2.47	0.36
	1.05	1.03	1	0	0.46	1.41	1	1.63	0.75	2.29	0.38
	1.4	1.38	1	0	0.46	1.34	1	1.55	0.75	2.18	0.4
	2.1	2.07	1	0	0.46	1.23	1	1.42	0.75	2	0.44
	3.50	3.45	1	0	0.46	1.10	1	1.27	0.75	1.79	0.49
	5.25	5.17	1	0	0.46	1.01	1	1.17	0.75	1.64	0.54
	7	6.89	1	0	0.46	1	1	1.16	0.75	1.63	0.54
α= 15°	0.178	0.172	1	0	0.44	1.47	1	1.65	0.72	2.39	0.38
	0.357	0.345	1	0	0.44	1.4	1	1.57	0.72	2.28	0.4
	0.714	0.689	1	0	0.44	1.3	1	1.46	0.72	2.11	0.43
	1.07	1.03	1	0	0.44	1.23	1	1.38	0.72	2	0.46
	1.43	1.38	1	0	0.44	1.19	1	1.34	0.72	1.93	0.47
	2.14	2.07	1	0	0.44	1.12	1	1.26	0.72	1.82	0.5
	3.57	3.45	1	0	0.44	1.02	1	1.14	0.72	1.66	0.55
	5.35	5.17	1	0	0.44	1	1	1.12	0.72	1.63	0.56
	7.14	6.89	1	0	0.44	1	1	1.12	0.72	1.63	0.56
α= 20°	. . .	. . .	1	0	0.43	1	1	1.09	0.7	1.63	0.57
α= 25°	. . .	. . .	1	0	0.41	0.87	1	0.92	0.67	1.41	0.68
α= 30°	. . .	. . .	1	0	0.39	0.76	1	0.78	0.63	1.24	0.8
α= 35°	. . .	. . .	1	0	0.37	0.66	1	0.66	0.6	1.07	0.95
α= 40°	. . .	. . .	1	0	0.35	0.57	1	0.55	0.57	0.93	1.14
α= 45°	. . .	. . .	1	0	0.33	0.5	1	0.47	0.54	0.81	1.34
			X	*Y*	*X*	*Y*	*X*	*Y*	*X*	*Y*	
Self-aligning ball bearings	. . .	. . .	1	0	0.4	0.4 cot α	1	0.42 cot α	0.65	0.65 cot α	1.5 tan α
Single-row radial contact separable ball bearings (magneto bearings)	. . .	. . .	1	0	0.5	2.5	. . .	. . .	. . .	. . .	0.2

(a) Permissible maximum value depends on bearing design internal clearance and raceway groove depth. Use first or second column depending on available information.

(b) Values of *X*, *Y*, and *e* for intermediate "relative axial loads" and α obtained by linear interpolation.

(c) For f_o values, see Ref 3.

Table 8 *X* and *Y* factors for radial roller bearings

Bearing type	$(F_a/F_r) \leq e$		$(F_a/F_r) > e$		e
	X	*Y*	*X*	*Y*	
Single row, $\alpha \neq 0$	1	0	0.4	0.4 cot α	1.5 tan α
Double row, $\alpha \neq 0$	1	0.45 cot α	0.67	0.67 cot α	1.5 tan α

Similarly, the dynamic equivalent load for a thrust bearing is that purely thrust load which, if applied to the bearing, would result in the same fatigue life as the actual applied load. Equation 14 can be used to determine the dynamic equivalent thrust load; however, *X* and *Y* factors are obtained from Tables 9 and 10 for thrust ball and roller bearings, respectively.

Table 9 *X* and *Y* factors for thrust ball bearings

α[a]	Single-direction bearings[b]		Double-direction bearings				e
	$(F_a/F_r) > e$		$(F_a/F_r) \leq e$		$(F_a/F_r) > e$		
	X	*Y*	*X*	*Y*	*X*	*Y*	
45°[c]	0.66	1	1.18	0.59	0.66	1	1.25
50°	0.73	1	1.37	0.57	0.73	1	1.49
55°	0.81	1	1.6	0.56	0.81	1	1.79
60°	0.92	1	1.9	0.55	0.92	1	2.17
65°	1.06	1	2.3	0.54	1.06	1	2.68
70°	1.28	1	2.9	0.53	1.28	1	3.43
75°	1.66	1	3.89	0.52	1.66	1	4.67
80°	4.8	1	5.86	0.52	2.43	1	7.09
85°	1.66	1	11.75	0.51	4.8	1	14.29
α = 90°	1.25 (tan α) $\left(1 - \frac{2}{3} \sin \alpha\right)$	1	$\frac{20}{13}$ (tan α) $\left(1 - \frac{1}{3} \sin \alpha\right)$	$\frac{10}{13}\left(1 - \frac{1}{3} \sin \alpha\right)$	1.25 (tan α) $\left(1 - \frac{2}{3} \sin \alpha\right)$	1	1.25 (tan α)

(a) Values of *X*, *Y*, and *e* for intermediate values of α are obtained by linear interpolation.

(b) $F_a/F_r < e$ is unsuitable for single-direction bearings.

(c) For thrust bearings $\alpha > 45°$. Values for $\alpha = 45°$ permit interpolation for α between 45° and 60°.

Table 10 *X* and *Y* factors for thrust roller bearings

Bearing type	$(F_a/F_r) \leq e$		$(F_a/F_r) > e$		e
	X	*Y*	*X*	*Y*	
Single direction, $\alpha \neq 90°$	[a]	[a]	tan α	1	1.5 tan α

(a) $F_a/F_r < e$ is unsuitable for single-direction bearings.

Nonstandard Application Conditions. The determination of bearing load rating and fatigue life using Eq 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 and the appropriate tables assumes that the bearing internal design and materials meet the limiting criteria and that operating conditions are not unusual; that is, they do not involve high speed, exceptionally heavy or light loading, high temperatures, and so on. Such conditions alter the bearing internal load distributions and contact stresses from those on which the load and life rating methods are based. For these situations, more sophisticated means of evaluation are available as detailed by Harris (Ref 10); these frequently require the use of a computer to perform the required analyses.

Fatigue Load Limit. Use of Eq 13 implies that even under extremely light loading, ball and roller bearing endurance is limited by rolling-contact fatigue. Ioannides and Harris (Ref 11), however, demonstrated that if a certain limiting stress is not exceeded in a given application, then rolling-contact fatigue can be eliminated as a consideration for bearing failure. This limiting stress is a function of the bearing internal design, the bearing material, and lubrication conditions. Use of a fatigue limit stress or endurance limit as a bearing selection criterion is demonstrated in Ref 12, which includes a fatigue load limit, P_u, for each bearing. Even if applied load exceeds the fatigue load limit, bearing fatigue endurance is longer than that predicted by Eq 13, depending on the magnitude of P/P_u. For $P \ggg P_u$, Eq 12 remains substantially accurate and is not conservative.

Effect of Contamination on Fatigue Life. In the selection of a bearing, the level of contamination in the application must be considered. Sayles and MacPherson (Ref 13) reported the effect of lubricant filtration on rolling-contact fatigue endurance. To obtain the fatigue lives predicted using Eq 13, especially for high-precision bearings manufactured from high-performance steels, it is necessary to minimize the contamination amount and particle size entering the bearing. Hard contaminants of sufficient size will dent the rolling-contact surfaces causing stress concentrations, interrupting lubricant film formation and reducing fatigue endurance. Hamer *et al.* (Ref 14) have shown that, depending on the application operating conditions, even soft contaminants can cause dents, thereby reducing endurance. Contamination can be precluded by providing a bearing with integral seals or by employing effective sealing means in the application. In the use of a fatigue load limit in the fatigue endurance calculation, Ref 12 takes into account the effect of contamination.

Basic Static Load Rating. According to Ref 3, 4, 5, experience shows that a total permanent deformation approximating 0.0001D at the center of the most heavily loaded rolling element-raceway contact can be tolerated in most bearing applications without impairing the operation of the bearing. Such indentation can occur due to loading when the bearing is not rotating or due to heavy shock load during operation. The basic static load rating is the load applied in such a way that it results in this permanent deformation magnitude. Furthermore, for most ball bearings, static loading that produces 4200 MPa (610 ksi) stress at the center of the most heavily loaded contact will also cause permanent deformation approximating 0.0001D. This equivalent stress is 4600 MPa (670 ksi) for self-aligning ball bearings and 4000 MPa (580 ksi) for roller bearings. Using the appropriate stress criterion, basic static load ratings are created for catalog ball and roller bearings.

For radial ball bearings, the basic static radial load rating (in newtons) can be determined from Eq 15:

$$C_{or} = f_o i Z D^2 \cos \alpha \qquad \textbf{(Eq 15)}$$

for which f_o values are given in Table 11. The values of Table 9 pertain to bearings having the same groove radii dimensional limitations as defined for basic radial load rating tabular data.

Table 11 f_o factor for ball bearings

Data based on the Hertz point contact formula with elastic modulus of 207 GPa (3.00 × 10^7 psi) and Poisson's ratio of 0.3

$(D \cdot \cos \alpha)/d$	Factor f_o [a]		
	Radial ball bearings[b]		Thrust ball bearings[c]
	Radial and angular-contact groove ball bearings	Self-aligning ball bearings	
0	14.7	1.9	61.6
0.01	14.9	2	60.8
0.02	15.1	2	59.9
0.03	15.3	2.1	59.1
0.04	15.5	2.1	58.3
0.05	15.7	2.1	57.5
0.06	15.9	2.2	56.7
0.07	16.1	2.2	55.9
0.08	16.3	2.3	55.1
0.09	16.5	2.3	54.3
0.10	16.4	2.4	53.5
0.11	16.1	2.4	52.7
0.12	15.9	2.4	51.9

0.13	15.6	2.5	51.2
0.14	15.4	2.5	50.4
0.15	15.2	2.6	49.6
0.16	14.9	2.6	48.8
0.17	14.7	2.7	48
0.18	14.4	2.7	47.3
0.19	14.2	2.8	46.5
0.20	14	2.8	45.7
0.21	13.7	2.8	45
0.22	13.5	2.9	44.2
0.23	13.2	2.9	43.5
0.24	13	3	42.7
0.25	12.8	3	41.9
0.26	12.5	3.1	41.2
0.27	12.3	3.1	40.5
0.28	12.1	3.2	39.7
0.29	11.8	3.2	39
0.30	11.6	3.3	38.2
0.31	11.4	3.3	37.5
0.32	11.2	3.4	36.8
0.33	10.9	3.4	36
0.34	10.7	3.5	35.3
0.35	10.5	3.5	34.6
0.36	10.3	3.6	. . .
0.37	10	3.6	. . .
0.38	9.8	3.7	. . .
0.39	9.6	3.8	. . .
0.40	9.4	3.8	. . .

(a) f_o values for intermediate values of $(D \cdot \cos \alpha)/d$ are obtained by linear interpolation.

(b) It is assumed that load distribution results in maximum ball load of $5F_r/(Z \cdot \cos \alpha)$.

(c) It is assumed that load distribution results in maximum ball load of $F_a/(Z \cdot \sin \alpha)$.

For thrust ball bearings, the basic static axial load rating (in newtons) is given by:

$$C_{oa} = f_o Z D^2 \sin \alpha \qquad \textbf{(Eq 16)}$$

for which f_o, values are also given in Table 11 and were Z is the number of balls carrying load in one direction. The basic static radial load rating for radial roller bearings is obtained from Eq 17:

$$C_{or} = 44 \left(1 - \frac{D \cdot \cos \alpha}{d}\right) iZlD \cdot \cos \alpha \qquad \textbf{(Eq 17)}$$

Similarly, the basic static axial load rating for thrust roller bearings is obtained from Eq 18:

$$C_{oa} = 220 \left(1 - \frac{D \cdot \cos \alpha}{d}\right) ZlD \cdot \sin \alpha \qquad \textbf{(Eq 18)}$$

where Z is the number of rollers carrying load in one direction. In cases where rollers have different lengths, Zl is the sum of the lengths of all rollers carrying load in a given direction.

As with dynamic loading, a static equivalent load must be defined. For a radial bearing, it is that purely radial static load which, if applied to the bearing, would produce the same permanent deformation as the applied static loading. For a radial bearing subjected to combined radial and thrust loading, Eq 19 gives the static equivalent radial load:

$$P_{or} = X_o F_r + Y_o F_a \qquad \textbf{(Eq 19)}$$

for which the factors X_o, and Y_o can be obtained from Tables 12 and 13 for radial ball and roller bearings respectively. If $F_r > P_{or}$ as calculated using Eq 19, then $P_{or} = F_r$ is used.

Table 12 X_o and Y_o factors for radial ball bearings

Bearing type	Single-row bearings		Double-row bearings	
	X_o	Y_o	X_o	Y_o
Radial contact groove ball bearings[(a)]	0.6	0.5	0.6	0.5
Angular-contact groove ball bearings, α= 15°	0.5	0.46	1	0.92
20°	0.5	0.42	1	0.84
25°	0.5	0.38	1	0.76
30°	0.5	0.33	1	0.66
35°	0.5	0.29	1	0.58
40°	0.5	0.26	1	0.52
45°	0.5	0.22	1	0.44
Self-aligning ball bearings, $\alpha \neq$ 0°	0.5	0.22 cot α	1	0.44 cot α

(a) Permissible maximum value of F_a/C_o depends on bearing design internal clearance and raceway groove depth.

Table 13 X_o and Y_o factors for radial roller bearings with α> 0°

Bearing type	X_o	Y_o
Single-row	0.5	0.22 cot α
Double-row	1	0.44 cot α

For thrust ball and roller bearings with α< 90°, static equivalent axial load is given by Eq 20:

$$P_{oa} = 2.3F_r \cdot \tan \alpha + F_a \quad \textbf{(Eq 20)}$$

Equation 20 is valid for all combinations of radial to axial load for double-direction bearings. It is strictly valid where $F_r/F_a \leq 0.44 \cot \alpha$; it gives satisfactory (but less conservative) results for $0.44 \cot \alpha < F_r/F_a < 0.67 \cot \alpha$. Thrust bearings with α= 90° can only support axial loading; in this case $P_{oa} = F_a$.

Frequently, loading applied to the bearing is more complex than that which has been considered using Eq 15, 16, 17, 18, 19, 20. In such cases, the specific load distributions corresponding to maximum permanent deformation of 0.0001*D,* or the equivalent stress, can be determined using methods given in Ref 10.

Rolling Bearing Friction

The principal sources of friction in roller bearings are due to sliding in the contact zones. Some friction occurs due to elastic hysteresis in rolling but the magnitude is insignificant compared to that caused by sliding motions. The magnitude of component sliding friction depends on the effectiveness of lubrication. For example, if adequately thick lubricant films are formed to effectively separate the contacting surfaces, friction forces will be relatively small. In the latter case, the friction torque of a rolling bearing will be substantially less than that of an oil-lubricated hydrodynamic or hydrostatic bearing in a given application.

The sources of sliding friction in a rolling bearing are as follows:

- Sliding in rolling element-raceway contacts due to geometry of contacting surfaces
- Sliding due to deformation of contacting elements (Heathcote slip)
- Sliding between the cage pockets and rolling elements

- In bearings with land-riding cages (see Fig. 17), sliding between the cage rails and ring-lands
- In roller bearings, sliding between roller ends and guiding inner-and/or outer-ring flanges
- For bearings having integral seals, sliding between the seal lips and the mating surfaces

Other than seal friction, which is substantially greater than all of the other friction sources, sliding in the rolling element-raceway contact areas is most significant in most applications, especially when the bearing is heavily loaded. When speeds are high, however, roller end-flange friction is very large in tapered roller bearings because the rollers are driven against the outer-ring (cup) flange with substantial force. Also, when speeds are high, depending on the amount of lubricant residing in the bearing cavity (free space), viscous drag friction of the lubricant through which the rolling elements orbit can be significant. In fact, excessive lubricant in the bearing cavity during high-speed operation can cause skidding (that is, gross sliding of the rolling elements over the inner raceway). Skidding in the absence of adequately thick elastohydrodynamic lubricant films can result in substantial damage of the contact surfaces and rapid bearing failure.

Bearing Friction Torque. To accurately predict the frictional torque of a ball or roller bearing in a given application is a complex process, considering all of the friction modes. Harris (Ref 10) provides details covering the analytical methods. In lieu of the more precise analytical methods, Palmgren (Ref 15) experimentally investigated bearing friction torque and established the following empirical formula:

$$M = M_l + M_v \qquad \textbf{(Eq 21)}$$

where M is the bearing total friction torque (in N · mm), M_l is the bearing friction torque (in N · mm) due to rolling element-raceway contact loading, and M_v is the bearing friction torque (in N · mm) due to hydrodynamic fluid friction.

M_v is independent of load, being a function of the viscosity and the amount of lubricant as well as the bearing speed. It is predominant when loading is light and speed significant. Equations 22 and 28define M_l and M_v, respectively:

$$M_l = f_1 F_\beta d_m \qquad \textbf{(Eq 22)}$$

where f_1 is a factor depending on bearing design and load, F_β is the applicable load (in N), and d_m is the bearing pitch diameter (in mm).

For ball bearings

$$f_1 = z\left(\frac{P_o}{C_o}\right)^y \qquad \textbf{(Eq 23)}$$

for which values of z and y can be obtained from Table 14 for various types of ball bearings. Table 15 gives values of f_1 for various roller bearing types with the exception of spherical roller bearings. Spherical roller bearings vary significantly in internal design, and therefore, the catalogs of manufacturers must be consulted for appropriate friction torque calculational methods. SKF uses the following formula:

$$M_l = f_1 F_\beta^a d^b \qquad \textbf{(Eq 24)}$$

for which f_1, F_β, a, and b values can be obtained from Ref 12for the different series of SKF spherical roller bearings.

Table 14 z and y factors for ball bearings

Bearing type	α(°)	z	y
Deep-groove	0	0.0002-0.0004[a]	0.55
Angular-contact	30, 40	0.001	0.33
Thrust	90	0.0008	0.33
Self-aligning	10	0.0003	0.4

(a) Lower values pertain to light-series bearings: higher values to heavy-series bearings.

Table 15 f_1 factors for roller bearing types

Bearing type	f_1
Cylindrical with cage	0.0002-0.0004[a]
Cylindrical, full complement	0.00055
Tapered	0.0004
Cylindrical roller thrust	0.0015
Spherical roller thrust	0.00023-0.0005[a]

(a) Lower values pertain to light-series bearings; higher values to heavy-series bearings.

Values of F_β can be obtained from the following equations. For ball bearings

$$F_\beta = 0.9F_a \cdot \cot \alpha - 0.1F_r \qquad \textbf{(Eq 25)}$$

If $F_r > F_\beta$, use $F_\beta = F_r$. For deep-groove ball bearings with 0° nominal contact angle, use

$$F_\beta = 3F_a - 0.1F_r \qquad \textbf{(Eq 26)}$$

For radial roller bearings:

$$F_\beta = 0.8F_a \cdot \cot \alpha \qquad \textbf{(Eq 27)}$$

Again, if $F_r > F_\beta$, use $F_\beta = F_r$. For either thrust ball or thrust roller bearings use $F_\beta = F_a$.

Equations 28 and 29are used to give estimates of viscous friction torque:

$$M_v = 10^{-7} f_v (vn)^{2/3} d^3 \quad vn > 2000 \qquad \textbf{(Eq 28)}$$

$$M_v = 160 \cdot 10^{-7} f_v d^3 \quad vn \leq 2000 \qquad \textbf{(Eq 29)}$$

where v is the kinematic viscosity of the lubricant (in cSt) at bearing operating temperature, and n is the bearing speed (in rev/min). Values of the factor f_v, a constant dependent on bearing type and lubrication, for different bearing types and various methods of lubrication are given in Table 16.

Table 16 f_v factors for bearing type and lubrication method

Bearing type	Type of lubrication			
	Grease	**Oil mist**	**Oil bath**	**Oil bath (vertical shaft) or oil jet**
Deep-groove ball[a]	0.7-2[b]	1	2	3-4[b]
Self-aligning ball[c]	1.5-2[b]	0.7-1[b]	1.5-2[b]	3-4[b]
Thrust ball	5.5	0.8	1.5	3
Angular-contact ball[a]	2	1.7	3.3	6.6
Cylindrical roller With cage[a]	0.6-1[b]	1.5-2.8[b]	2.2-4[b]	2.2-4[b] [d]
Full complement	5-10[b]	. . .	5	. . .
Spherical roller[c]	3.5-7[b]	1.7-3.5[b]	3.5-7[b]	7-14[b]
Tapered roller[a]	6	3	6	8-10[b] [d]
Needle roller	12	6	12	24
Thrust cylindrical roller	9	. . .	3.5	8
Thrust spherical roller	. . .	. . .	2.5-5[b]	5-10[b]
Thrust needle roller	14	. . .	5	11

(a) Use $2 \cdot f_v$ value for paired bearings or double-row bearings.

(b) Lower values pertain to light-series bearings; higher values to heavy-series bearings.

(c) Double-row bearings only.

(d) For oil bath lubrication and vertical shaft, use $2 \cdot f_v$.

Radial cylindrical roller bearings with flanges on both inner and outer rings can carry thrust load in addition to radial load. In this situation, the roller ends are loaded against one flange on each ring. Friction torque due to roller end-flange sliding depends on specific design of the roller end and flange geometries. In most cases, the roller corners bear on flanges that are angled a few minutes from radial planes. References 10and 12 can be consulted for further details regarding calculation of friction torque due to roller end-flange sliding under thrust loading.

Wear

Wear (in regard to rolling-element bearings) is the loss or displacement of material from any of the rolling bearing internal contact surfaces. Material loss can be in the form of loose debris, whereas material displacement can occur by local plastic deformation or transfer of material from one location to another by adhesion. When wear has occurred to the extent it threatens the essential bearing function, the bearing is considered failed. Bearing engineers recognize the wear modes listed below as contributing to bearing failure:

- Mild mechanical wear
- Adhesive wear
- Smearing
- Corrosive wear
- Plastic flow
- Surface indentation
- Abrasive wear
- Surface distress
- Pitting
- Fatigue spalling

These failure modes are defined without presupposing the exact mechanisms that cause them. They are defined in engineering terms based on observations of the evidence remaining from complex sequences of events involving many physical and chemical processes. Some of these occurred in the manufacturing of the original surfaces. Associated with

these physical/chemical interactions are several mechanistic wear processes (for example, adhesion, chemical reaction, plastic flow, and fatigue). These wear processes are described in the Sections on wear in this Handbook.

Recognized wear modes in ball and roller bearings (for example, smearing and pitting) are not uniquely associated with any of the wear processes but rather with their interactions. Attempts are made in the following discussion to connect the wear process to the failure mode to best determine how to eliminate or minimize wear through improved lubrication, alternate material selection, and application design.

In rolling bearings, smearing (also called galling or scuffing) is adhesive wear on a macroscopic scale. Smearing occurs when sliding is substantial and lubricant film thicknesses in the rolling contacts are insufficient to adequately separate the rolling bodies. It is accompanied by severe plastic deformation, and it involves the gross failure of the surface, increased friction, and increased contact temperatures. These phenomena can lead to rapid bearing failure. Figure 28 is an example of smearing on a cylindrical roller surface.

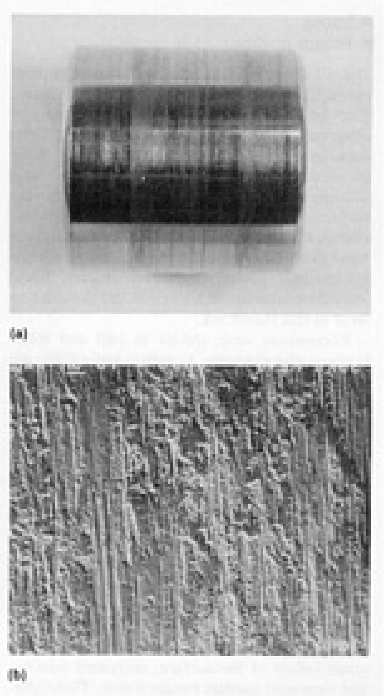


Fig. 28 Smearing (severe mechanical wear) on the surface of a cylindrical roller shown at two magnifications. (a) 4×. (b) 410×

As indicated above, if a rolling bearing is overloaded while not rotating, plastic deformation (Brinelling or denting) of the rolling surfaces can occur that leads to noisy operation and possible failure. Denting can also occur as a result of overrolling solid contaminants. Once the wear process has begun, these phenomena can be augmented by wear debris. Figure 29 illustrates hard-particle contaminant denting in a raceway surface. Plowing occurs due to displacement of bearing material by a hard particle in the presence of sliding. Figure 30 depicts such damage in a bearing ball surface. Abrasive wear, such as plowing, sometimes occurs when hard surfaces interact with softer surfaces in the presence of a contaminant; the softer cage materials are especially susceptible.

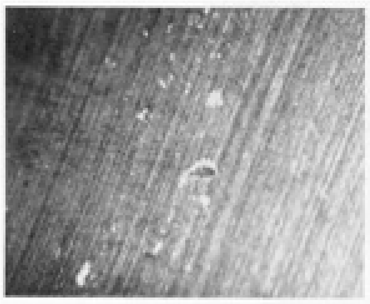


Fig. 29 Ball bearing raceway denting caused by hard particle contaminant. 15×

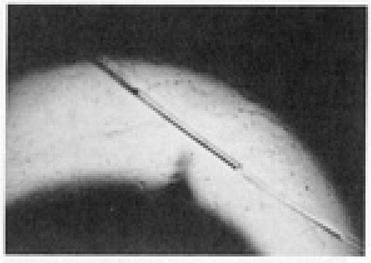


Fig. 30 Plowing damage generated on the surface of a bearing ball. 11×

Rolling-contact fatigue has been described in the section "Load Ratings" in this article. When loading is heavy and after a sufficient number of load cycles has been accumulated, spalling can occur (Fig. 22). In a well-designed and well-lubricated bearing application, the maximum shear stresses occur below the rolling-contact surfaces. In such cases, a spall is a pit that commences as a crack below the rolling-contact surface. After repeated over-rollings, the crack propagates and some material is expelled. In the presence of substantial sliding (in addition to rolling), the maximum shear stresses move closer to the surface and pitting can occur directly at the surface (Fig. 31).

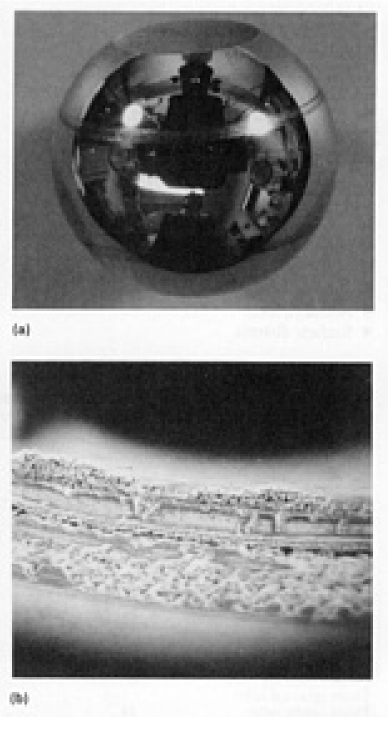


Fig. 31 Pitting (surface distress) generated on a rolling-contact surface. (a) Ball used in V-groove testing machine. (b) Pit produced on ball. 28×

Wear Control

Wear control in rolling bearings is accomplished by effective separation of contacting surfaces by lubricant films generated by hydrodynamic and elastohydrodynamic fluid action. In the absence of such processes, wear will occur in varying degrees and rates depending on the effectiveness of boundary lubrication and solid-film lubrication schemes. Numerous methods exist for increasing wear performance (for example, utilizing additives in lubricants, changing material pairs, modifying contacting surfaces, and so on). Unless effective surface separation is achieved, however, wear will occur. In general, significant wear does not occur during the operating lifetime of well-designed and manufactured ball and roller bearings that are properly lubricated with viscous oil-type fluids. Therefore, most rolling bearing manufacturers do not provide information regarding the wear rate to be expected in properly designed applications. An exception to this is Ref 16. However, in using this reference it should not be presumed that the data apply to the rolling bearing products of other manufacturers.

Lubricant Film Thickness. To avoid bearing wear in a given application, it is necessary for the rolling-contact surfaces to be almost completely separated by lubricant films. Using the method of Hamrock and Dowson (Ref 17), the minimum lubricant film thickness can be estimated for rolling bearing point contacts:

$$H = \frac{2.69\bar{U}^{0.67}G^{0.53}(1 - 0.61e^{-0.73\kappa})}{\bar{Q}_e^{0.067}} \quad \textbf{(Eq 30)}$$

in which:

$$H = \frac{h}{R_e} \quad \textbf{(Eq 31)}$$

where h is the minimum lubricant film thickness (in mm) and R_e is the equivalent radius of curvature (in mm), and in which:

$$R_e = \frac{1}{\dfrac{1}{r_I} + \dfrac{1}{r_{II}}} \quad \textbf{(Eq 32)}$$

$$\bar{U} = \frac{\eta_0 U}{2\,E'R_e} \quad \textbf{(Eq 33)}$$

$$G = \lambda E' \quad \textbf{(Eq 34)}$$

$$E' = \frac{E}{1 - \xi^2} \quad \textbf{(Eq 35)}$$

$$\bar{Q}_e = \frac{Q}{E'R_{ex}R_{ey}} \quad \textbf{(Eq 36)}$$

where r_I is the radius of rolling body I (in mm), r_{II} is the radius of rolling body II (in mm), E is the modulus of elasticity (in MPa), ξ is Poisson's ratio $\overline{U}$ is the rolling velocity (dimensionless), η_0 is the absolute viscosity at contact inlet (in N · s/mm^2), λ is the pressure coefficient of viscosity in MPa^{-1}, and $\overline{Q}_e$ is the contact stress (dimensionless); and in which:

$$\kappa = \frac{a}{b} \qquad \textbf{(Eq 37)}$$

where κ is the ellipticity ratio of the contact area, a is the semimajor axis of the elliptical contact area (in mm), and b is the semiminor axis of the elliptical contact area (in mm); and in which:

$$U = \frac{u_I + u_{II}}{2} \qquad \textbf{(Eq 38)}$$

where u_I is the surface velocity of body I in the rolling direction (in mm/s), and u_{II} is the surface velocity of body II in the rolling direction (in mm/s).

Equation 30 pertains to Fig. 32, which shows two bodies in rolling contact. To use Eq 30 for a rolling bearing, it is necessary to determine the distribution of the applied load among the rolling elements and also the speeds of each of the rolling components. This is a rather complex process, and rigorous analytical methods of evaluation are given in Ref 10. Once the load distribution has been determined, it is only necessary to establish the adequacy of the lubricant thickness at the maximum-loaded rolling element to determine whether the bearing is lubricated sufficiently. In lieu of the rigorous calculation method, for simple radially loaded bearings, the maximum rolling-element load can be approximated by:

$$Q_{max} = \frac{5F_r}{iZ \cdot \cos\alpha} \qquad \textbf{(Eq 39)}$$

For a simple thrust-loaded bearing:

$$Q_{max} = \frac{F_a}{Z \cdot \sin\alpha} \qquad \textbf{(Eq 40)}$$

Bearing internal speeds can be estimated from the following equations:

$$n_m = \frac{[n_i(1 - \gamma) + n_o(1 + \gamma)]}{2} \qquad \textbf{(Eq 41)}$$

$$n_{RE} = \frac{d}{2D}(1 - \gamma)(1 + \gamma)(n_o - n_i) \qquad \textbf{(Eq 42)}$$

where n_m is the cage speed (rolling-element orbital speed), n_i is the inner ring speed, n_o is the outer ring speed, and n_{RE} is the ball or roller speed about its own axis.

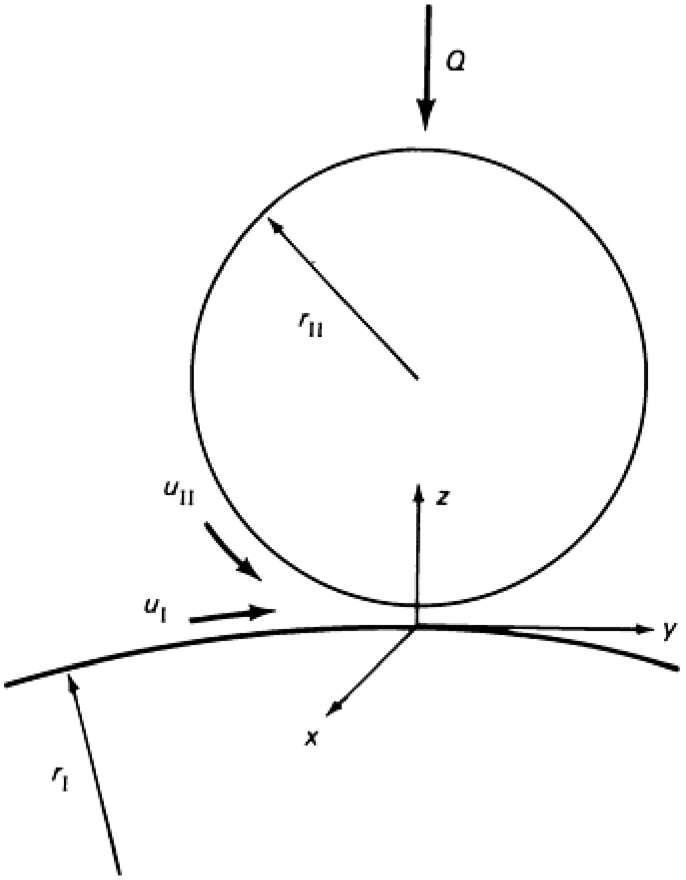


Fig. 32 Measurable parameters present when two bodies of radius r_I and r_{II} are in rolling contact with each other

Using the rolling-element and bearing ring radial dimensions, it is possible to estimate average surface velocities u_I and u_{II}. Surface velocities can also be determined in a more precise manner (Ref 10); however, for many applications, the approximate calculation method will suffice.

To determine whether the rolling elements are separated adequately from the raceway surfaces by lubricant films, it is necessary to compare the calculated lubricant film thickness to the composite roughness of the mating surfaces. The rms value is used such that:

$$\sigma_e = \sqrt{\sigma_I^2 + \sigma_{II}^2} \qquad \textbf{(Eq 43)}$$

where σ_I is the surface roughness of body I and σ_{II} is the surface roughness of body II; and:

$$\Lambda = \frac{h}{\sigma_e} \qquad \textbf{(Eq 44)}$$

In the section "Bearing Fatigue Life" of this article, it was indicated that when lubrication conditions are very favorable and contamination is minimal, extended bearing rolling-contact fatigue lives can be obtained in an application. This means that virtually complete separation of rolling elements from raceways by lubricant films has been achieved and frictional heating is effectively accommodated. In this situation, wear of rolling-contact surfaces has also been virtually eliminated.

When lubricant film thicknesses are insufficient to obtain complete separation, fatigue life can still be satisfactory, provided that contamination is minimal and gross sliding in the contacts does not occur. Microslip in the contacts will still occur owing to contact deformations that preclude achievement of pure rolling. However, in the absence of gross sliding and contamination, for $0.4 \leq \Lambda \leq 1$, wear will generally not be significant. If either of these conditions is prevalent, distress of the rolling-contact surfaces can be anticipated. This distress can lead to early bearing fatigue failure or thermal imbalance and ultimate bearing seizure. Contamination can be prevented from entering the bearing by effective filtration of the lubricant, by sealing the bearing, or by both. Gross sliding is a function of bearing design and operating conditions, and it tends to occur in high-speed applications when the bearing applied load is relatively light. Harris (Ref 10) demonstrates analytical methods to predict whether gross sliding (skidding) will occur during steady-state operation. In such cases, preloading of the bearing or of selected rolling elements can be used to minimize the amount of sliding. Gross sliding can also occur, however, during rapid accelerations and decelerations of bearing speed or rolling-element speeds. For example, smearing damage can occur during slow-speed operation of radially loaded large roller bearings.

In addition to achieving adequate separation of rolling-contact surfaces, effective lubrication of cage-rolling element and cage rail-ringland sliding surfaces must be achieved. In the case of roller bearings, adequate lubrication of the roller end-guide flange contacts must be provided. Insufficient lubrication of these principally sliding contact surfaces can result in excessive friction and wear, particularly when contamination is present. Excessive roller end-flange friction can result in roller skewing and rapid bearing failure. Furthermore, wear debris from such inadequately lubricated surfaces can enter the bearing, subsequently affecting rolling-contact surface distress and bearing failure. To ensure adequate lubricant flow, the lubrication system and bearing design can be arranged to direct the lubricant to the sliding surfaces. For example, oil flow can be directed first through passages under the bearing and then through radial holes drilled between the bearing bore and the inner raceway undercuts to ensure roller end-flange lubrication in cylindrical and tapered roller bearings. Oil jets can be directed to impinge on cage rail-ringland interfaces. In a similar way, critical sliding surfaces can be lubricated to prevent excessive friction and wear. In all cases, minimization of contamination is essential in order to bring about minimization of friction and wear.

References

1. D. Dowson, *History of Tribology,* Longman, 1979
2. "Rolling Bearings--Dynamic Load Ratings and Rating Life," ISO 281-1, International Standards Organization, 1991
3. "Rolling Bearings--Static Load Ratings," ISO 76, International Standards Organization, 1987
4. "Load Ratings and Fatigue Life for Ball Bearings," ANSI/AFBMA Std 9-1990, American National Standards Institute/Antifriction Bearing Manufacturers' Association, 1990
5. "Load Ratings and Fatigue Life for Roller Bearings," ANSI/AFBMA Std 11-1990, American National Standards Institute/Antifriction Bearing Manufacturers' Association, 1990
6. W. Weibull, A Statistical Theory of the Strength of Materials, *Proceedings of the Royal Swedish Institute of Engineering Research* (Stockholm), 1939, p 151
7. G. Lundberg and A. Palmgren, Dynamic Capacity of Rolling Bearings, *Acta Polytechnica, Mech. Eng. Series,* Vol 1 (No. 3), Royal Swedish Academy of Engineering Sciences, 1947, p 7
8. G. Lundberg and A. Palmgren, Dynamic Capacity of Roller Bearings, *Acta Polytechnica, Mech. Eng. Series,* Vol 2 (No. 4), Royal Swedish Academy of Engineering Sciences, 1952, p 96
9. G. Lundberg, A. Palmgren, and E. Bratt, Statiska Bärformågan hos Kullager och Rullager, *Kullagertidningen,* No. 3, 1943
10. T. Harris, *Rolling Bearing Analysis,* 3rd ed., Wiley, 1991
11. E. Ioannides and T. Harris, A New Fatigue Life Model for Rolling Bearings, *J. Tribol. (Trans. ASME),* Vol 107, July 1985, p 367-378
12. SKF, General Catalogue 4000E, 1989
13. R. Sayles and P. MacPherson, Influence of Wear Debris on Rolling Contact Fatigue, *Rolling Contact Fatigue Testing of Bearing Steels,* STP 771, J. Hoo, Ed., American Society for Testing and Materials, 1982, p 255-274
14. J. Hamer, R. Sayles, and E. Ioannides, Particle Deformation and Counterface Damage when Relatively Soft Particles Are Squashed between Hard Anvils, *STLE Tribology Trans.,* Vol 32 (No. 3), 1989, p 281-288

15. A. Palmgren, *Ball and Roller Bearing Engineering,* 3rd ed., Burbank, 1959
16. P. Eschmann, L. Hasbargen, and K. Weigand, *Ball and Roller Bearings, Theory, Design and Application,* John Wiley & Sons, 1985, p 188-194
17. B. Hamrock and D. Dowson, Isothermal Elastohydrodynamic Lubrication of Point Contacts--Part III--Fully Flooded Results, *J. Lubr. Technol. (Trans. ASME),* Vol 99, 1977, p 264-276

Friction and Wear of Sliding Bearings

Ron Pike and J.M. Conway-Jones, Glacier Vandervell Inc.

Introduction

SLIDING BEARINGS (plain bearings) are designed to transmit force between two surfaces that are in relative motion. The simplest forms of sliding bearings are used unlubricated and thus suffer from the penalty of high friction and wear. Providing lubricant under favorable conditions separates the surfaces and reduces the coefficient of friction by a factor of $\geq$100, and reduces the rate of wear by many orders of magnitude. In order to optimize the performance of sliding bearings, a variety of materials have been used (see the article "Friction and Wear of Sliding Bearing Materials" in this Volume).

A bearing may be an integral part of the equipment, but it is usually a separate component either in the form of a round bushing, a half bearing, a thrust washer, a flanged bearing (which can accommodate both radial and axial loads), or a wear plate. These components are generally classified as thin wall if their thickness is $\leq$5 mm ($\leq$0.2 in.) and heavy wall if their thickness is >5 mm (>0.2 in.). Bearings with diameters >150 mm (>6 in.) are mostly in the heavy-wall class. For special-purpose applications, the bearing surface may be divided into a number of elements or tilting pads, either for relatively highly loaded high-speed thrust bearings, or for lightly loaded journal bearings where the increased resistance to oil-film whirl provided by these bearings is required (as in vertical-shaft high-speed pumps, turbo compressors, and so on).

Sliding bearing performance is substantially affected by the lubrication regime. Three categories of lubrication regimes will be discussed in this article:

- Rubbing (dry or boundary lubrication)
- Full film (surfaces separated by lubricant film)
- Mixed film (load partly carried by rubbing and partly by fluid film)

Effect of *PV* Factor on Rubbing Bearing Performance

The potential for wear and friction depends on the intensity of contact between the surfaces. Heat generated through rubbing contact depends upon the load, the speed, and the coefficient of friction (μ). Because the first two parameters can vary over many orders of magnitude, it is common to categorize the severity of applications in terms of the *PV* factor (Fig. 1):

$$\frac{\text{Work}}{\text{Area}} = \frac{\mu WV}{A} \quad \textbf{(Eq 1)}$$

$$\text{Heat generated } \alpha PV \quad \textbf{(Eq 2)}$$

where *P* is the specified load or unit load; *W* is the load; *A* is the projected area ($A = d \times b$, where *d* is the diameter and *b* is the width for a journal bearing); and *V* is the rubbing velocity.

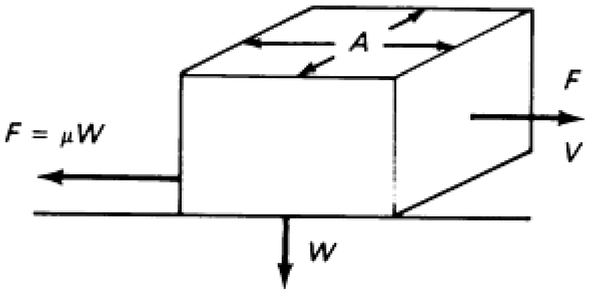


Fig. 1 Schematic showing key components required to calculate work done against friction. *F,* frictional force; μ , coefficient of friction; *W,* load; *V,* velocity; *A,* area of rectangle

The coefficient of friction will depend on the material composition of the surfaces, the surface finish, the temperature, and the effectiveness of the lubrication. For a dry bearing (no fluid lubricant used), any heat, *Q,* generated by friction must be dissipated by conduction, and the temperature rise of the surface, T_s, above that of the environment, T_E, can be related to the *PV* factor by (Fig. 2):

$$Q = \frac{kA'(T_s - T_E)}{t} = \mu PVA \quad \textbf{(Eq 3)}$$

$$(PV)_{max} \propto \frac{k}{\mu t}\left(\frac{A'}{A}\right)(T_s - T_E)_{max} \quad \textbf{(Eq 4)}$$

where *k* is the thermal conductivity; *t* is the material thickness; and *A*' is the area shown in Fig. 2.

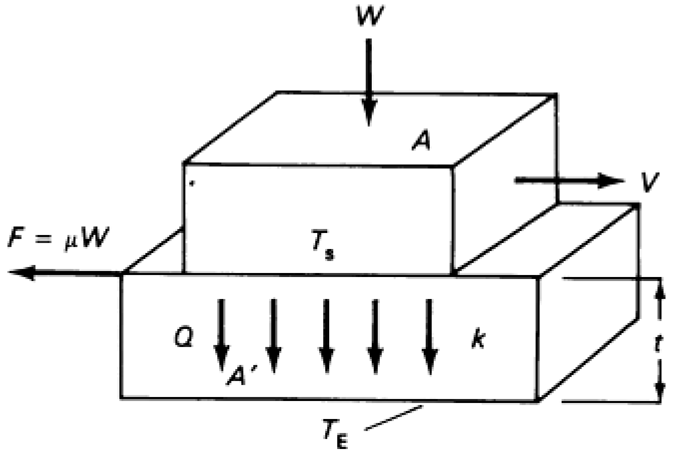


Fig. 2 Schematic showing parameters that affect frictional heat dissipated by conduction. *Q,* heat generated by friction; *A*', area of larger rectangle; *k,* thermal conductivity; T_s, surface temperature; T_E, temperature of environment

Thus, for rubbing contacts at high specific loads or high velocities, the bearing materials are required to have a high thermal conductivity, a low coefficient of friction, and the ability to accept high surface temperatures. If the bearing material has a low thermal conductivity (for example, polymers), it is preferable that the thickness is very small; otherwise the permissible *PV* factor will be correspondingly limited.

Wear Rate for Dry Rubbing

After an initial break-in (bedding-in), the rate of dry wear, *H/t,* for many materials becomes constant (that is, proportional to the *PV* factor). Thus:

$$H/t = K \cdot PV \qquad \textbf{(Eq 5)}$$

Table 1 defines the terms in Eq 5 in both English and metric units.

Table 1 Nomenclature and units for calculating specific wear rate of sliding bearings

Parameter	Symbol	SI units	English units
Depth of wear	*H*	m	in.
Time	*t*	s	min
Specific wear rate	*K*	$m^3/N \cdot m$	$in.^3/lbf \cdot ft$
Specific load	*P*	N/m^2	psi
Rubbing velocity	*V*	m/s	ft/min

The specific wear rate is the volumetric wear divided by the load and distance rubbed. The allowable specific wear rate is:

$$K_a = H_a/PVt_r \qquad \textbf{(Eq 6)}$$

where H_a is the allowable depth of wear and t_r is the required life. The specific wear rate also increases with the surface temperature and the surface roughness and can also depend upon the material of the mating surface.

Metallic Bearings. For metals with high thermal conductivity, the preferred method of improving *PV* capability and reducing the wear rate is to reduce the coefficient of friction by the addition of lubricants. For low speeds (typically, <0.5 m/s, or 100 ft/min), this may take the form of greases or solid lubricants such as graphite. At high speeds, oil can be used if provision is made to recirculate the oil. Thus, the bearing can be made porous, such that the oil is retained by capillary action (for example, porous bronze bearings used in small electric motors with typical shaft sizes of $\leq$6 mm, or $\frac{1}{4}$ in.). In larger electric motors (for use with bearings that accommodate shaft diameters to $\leq$15 mm, or $\frac{5}{8}$ in.), babbitt-lined bearings with wick lubrication can be used. The bearing housing is packed with cotton waste or a proprietary cellulose material that is soaked in oil and that collects the oil expelled form the end of the bearing and recirculates it to a large hole in the unloaded surface by capillary action.

Table 2 indicates typical coefficient of friction values and permissible *PV* factors for bearings that rely on conduction for heat dissipation. Depending upon the compatibility of the surfaces (for information about mutual polishing and nonscoring surfaces, refer to the article "Friction and Wear of Sliding Bearing Materials" in this Volume), the coefficient of friction may decrease after the initial break-in period but will increase towards the end of the useful life of the bearing. This increase in friction may be accompanied by an increase in the rate of wear, in scoring, and so on, which may be related to the deterioration of the lubricant or loss of the lubricant by evaporation, and so on. The life of greased bearings can often be enhanced by regreasing at regular intervals. Providing indentations or large grooves in the surface to retain the grease is also beneficial.

Table 2 Typical parameters for selected metallic and polymer bearings rubbing against ground steel shaft with R_a of 0.4 μm

Bearing Material	Lubrication	Coefficient of friction[a]	Permissible *PV*		Specific wear rate, *K*[b]
			MPa · (m/s)	**[psi · (ft/min)] × 10^3**	
Cast iron and lead bronze	Dry	0.2-0.5	0.1	3	10-1000
Cast iron and lead bronze	Greased	0.05-0.15	1	30	0.01-1
Porous bronze, graphite impregnated	Dry	0.1-0.4	0.2	6	0.5-5
Porous bronze	Oil impregnated	0.01-0.1	1	30	0.001-0.01
Babbitt metal (steel backed)	Wick lubricated	0.005-0.1	2	60	0.001-0.01
Polytetrafluoroethylene (PTFE), 3 mm ($\frac{1}{8}$ in.) thick	Dry	0.05-0.3	0.05	1.5	10-1000
Steel-backed porous bronze impregnated with PTFE and lead	Dry	0.05-0.3	1	30	0.03-0.3
Polyoxymethylene (POM) 0.2 mm (0.008 in.) thick bonded to porous bronze with steel backing	Indentations greased during assembly	0.05-0.15	1.5	45	0.003-0.03

(a) For dry rubbing, the coefficient of friction will generally be higher at low loads, high speeds, high temperature, and for rougher surface finishes. Except for PTFE, the static (breakaway) coefficient of friction is 30 to 50% higher than the dynamic coefficient. For lubricated rubbing, the coefficient of friction may be higher at high loads.

(b) For SI units, multiply *K* by 10^{-15} to obtain $m^3/N \cdot m$; for customary units, multiply *K* by 0.8×10^{-10} to obtain $in.^3/lbf \cdot ft$.

Polymer Bearings. Although polymer materials have poor thermal conductivity, experience has shown that if used in the form of a thin layer bonded to a steel backing (generally through a layer of sintered porous bronze), polymer materials can provide adequate performance for a wide range of applications; these can operate dry or with initial lubrication only. Dry polymer material applications include car door hinges, office machinery, heat-sealing packaging machines, pneumatic equipment, and aerospace applications. Because of the wide range of contributing factors, it is not possible to quote rates of wear except in very specific circumstances. Successful operation is based largely on previous experience or simulated test conditions. Table 2 can provide only a general indication of the specific wear rates. Many proprietary dry bearing formulations have been developed, and further details of the performance of these materials can be obtained from manufacturer catalogues or from Ref 1.

Full-Film Lubricated Bearings

The presence of oil between rubbing surfaces not only reduces the coefficient of friction but, if supplied continuously and in sufficient quantity, can act as an additional method of heat removal. Under favorable conditions, sufficient oil can be provided to completely separate the rubbing surfaces, thereby minimizing wear and, principle, allowing the bearing to operate for many thousands or even hundreds of thousands of hours. In fact, it is not possible to quote a rate of wear. For such bearings, wear only occurs during starting and stopping or as a result of deterioration or loss of the lubricant, contamination by abrasive particles, and so on. The bearing life may be terminated by other conditions such as corrosion or, for a dynamically loaded bearing, by cavitation erosion or fatigue. These factors will be considered in later sections of this article. Additional information is available in the Section "Wear and Surface Damage Overview" in this Volume.

The maintenance of a full film separating surfaces requires:

- Adequate and continuous supply of lubricant
- Speed to drag sufficient oil into the loaded area
- Convergence between the bearing and mating surface such that pressure is built up in the lubricant film to oppose the load

This is known as hydrodynamic lubrication and relies on the viscosity of the oil to provide the oil's resistance to escape, thereby generating high oil film pressure between the closely converged surfaces. The convergence is achieved by

providing a clearance between the journal and the bearing, typically 0.001 mm/mm of diameter. Under the influence of load, the journal moves to an eccentric position within the bearing, thereby causing convergence of the oil film and generating pressure that balances the load. For a flat surface, such as a thrust washer, if specific loads >1 MPa (>150 psi) are to be supported, it is advisable to machine a series of tapered lands on the surface with a taper of 0.001 mm/mm. Alternatively, a number of tilting pads can be arranged with crowned surfaces (or flat with offset pivots) such that the pads tilt to form the necessary convergence of the oil film. For a journal bearing, as the load increases the journal moves to greater eccentricity such that the minimum oil film thickness, h_{min}, decreases.

In hydrodynamic lubrication theory, the relationships among the variables have been expressed in a series of equations, all of which involve the dimensionless group $\eta N/P$ multiplied by the ratio $(d/c_d)^2$, which is a form of the Sommerfeld number and where d is the shaft diameter and c_d is the diametrical clearance. (Additional information can be found in Ref 2.) Various equations, computations, and graphical representations have been established for the relationship between the ratio of minimum oil film thickness to radial clearance (h_{min}/c_r) and the Sommerfeld number. Although these parameters are useful if the viscosity is known, it will be recognized that as the speed increases and the rate of shear increases, heat will be generated within the oil and the effective viscosity will be reduced. It is necessary to carry out a heat balance in order to establish the appropriate value of the Sommerfeld number. A rigorous method of doing this for steadily loaded bearings has been published in Ref 3, and a series of design charts have been prepared by F.A. Martin and D.R. Garner (Ref 4). Figure 3 is based on Ref 4 with provision made for the specification of oil by ISO viscosity grade (centistokes, cSt, at 40 °C, or 105 °F). In using this chart, it should be noted that it is not necessary to estimate the effective operating viscosity because this has already been derived during the preparation of the chart (assuming oil is supplied at 50 °C, or 122 °F).

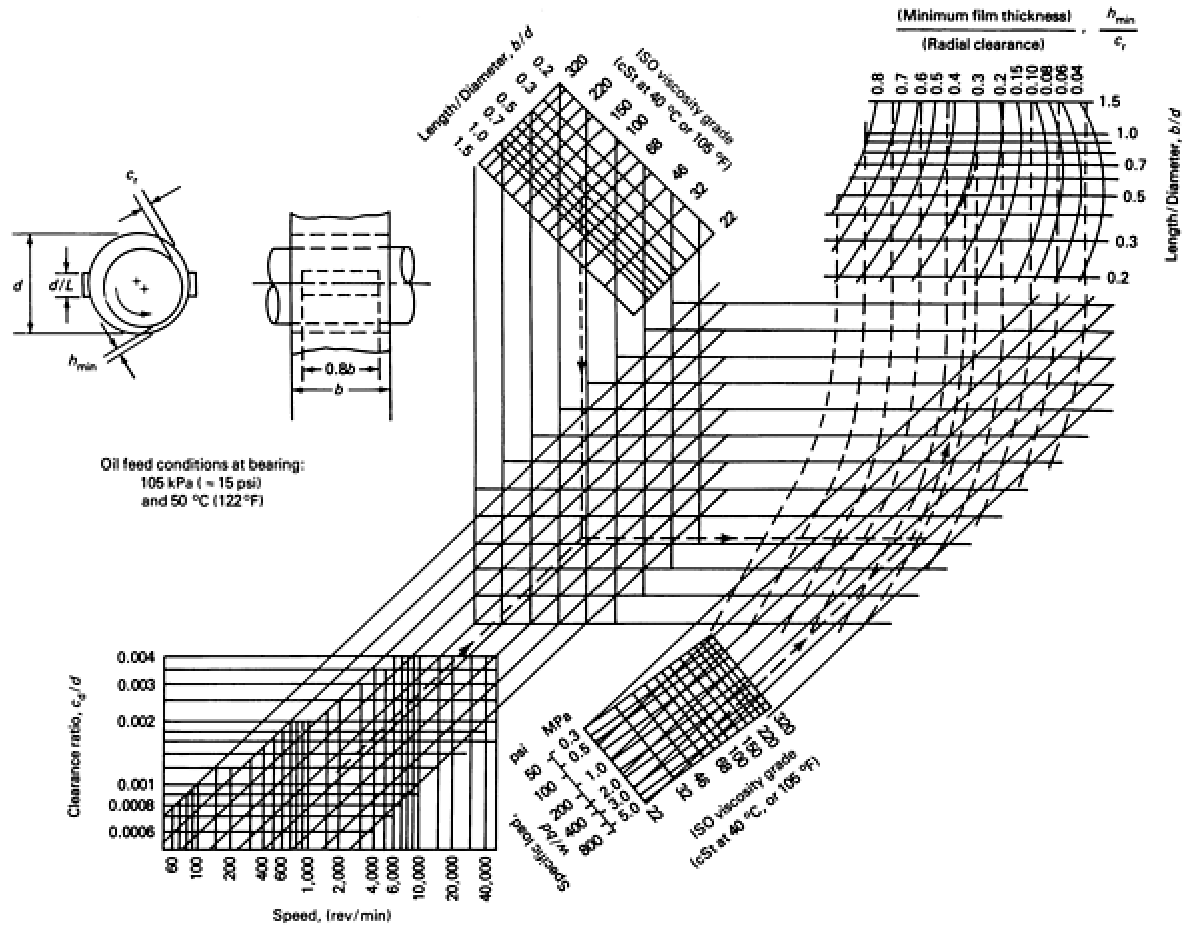


Fig. 3 Nomographs used to predict minimum oil film thickness for steadily loaded bearings

Permissible Minimum Oil Film Thickness. Typical values of h_{min}/c_r for steadily loaded hydrodynamic bearings are in the range of 0.4 to 0.2. For higher ratios at higher speed (typically >3000 rev/min), oil film whirl is a possibility (Ref 2) and a bearing with noncircular bore or tilting pads may be required. If the value of the ratio is <0.2, there is the possibility

of asperity contact, which can lead progressively to increased friction and wear, overheating, and possibly seizure (particularly if some misalignment, waviness, or other local geometric deviations are present from faulty assembly, and so on).

Depending on the selection of bearing materials and the operating conditions, it may be possible for the bearing to break-in (bed-in) during initial running and develop a geometry that conforms more closely with the alignment of the journal. Furthermore, if a bearing material is chosen that is compatible with the journal, it is possible that some asperity polishing will occur, reducing the sum of the initial heights of the asperities on the journal and the bearing so that full hydrodynamic separation is established for $h_{min}/c_r \leq 0.1$.

Effect of Asperity Height (Shaft Roughness). Typically, the minimum permissible oil film thickness calculated from hydrodynamic theory for steadily loaded journal bearings is in the range of 0.0025 to 0.025 mm (0.0001 to 0.001 in.). This can be related to the sum of the maximum heights of the asperities on the shaft (e_s) and bearing (e_B) by the Lambda ratio (Fig. 4):

$$\lambda = h_{min}/(e_s + e_B) \qquad \textbf{(Eq 7)}$$

Instead of basing the value of λ on the arithmetic sum of asperities, some engineers prefer to use the statistical sum $\sqrt{(e_S^2 + e_B^2)}$ instead of the arithmetic sum shown in Eq 7. The statistical sum may be appropriate for static or rolling contact, but the implied interlocking of the surface profiles is not appropriate for surfaces that are moving past each other; for these surfaces, the arithmetic sum is more relevant.

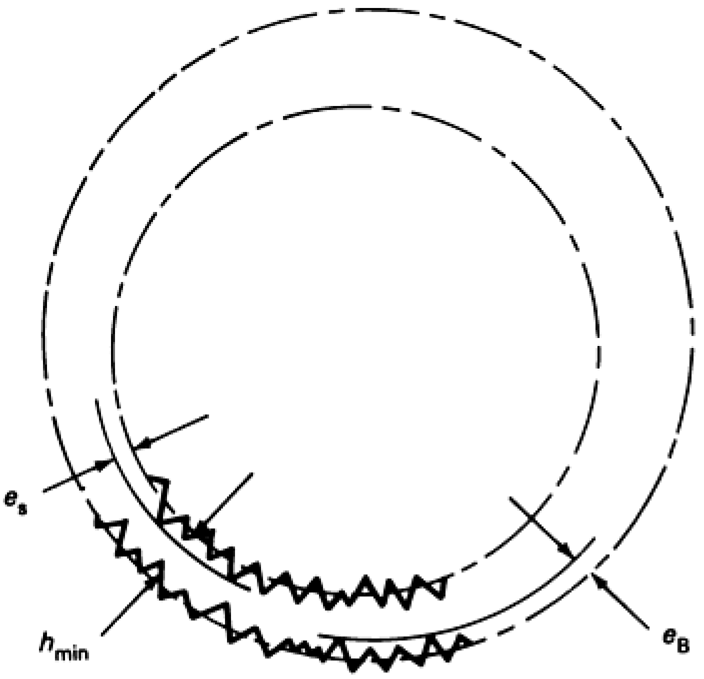


Fig. 4 Typical surface asperities on a bearing and its mating shaft

Of primary importance is the appropriate parameter for defining the maximum height of the asperities above the average level of the surface. The single predominant peak height (R_p), or "leveling depth" (Ref 5), may not be relevant because it can easily be rubbed off. A more typical value is the mean height of the highest peaks on five adjacent sampling lengths, R_{pm} (Fig. 5a). For a ground surface with normal (Gaussian) distribution of asperity heights, this is about four times the arithmetic average height above the mean line, R_a (or AA) (Fig. 5b), or about 3σ, where σ is the standard deviation or

root-mean-square roughness, R_q. If e_s and e_B are based on R_{pm} values of surface finish, the surfaces will be separated by the oil film if $\lambda R_{pm} > 1$. Some authorities base e_s and e_B on σ, in which case $\lambda\sigma > 3$.

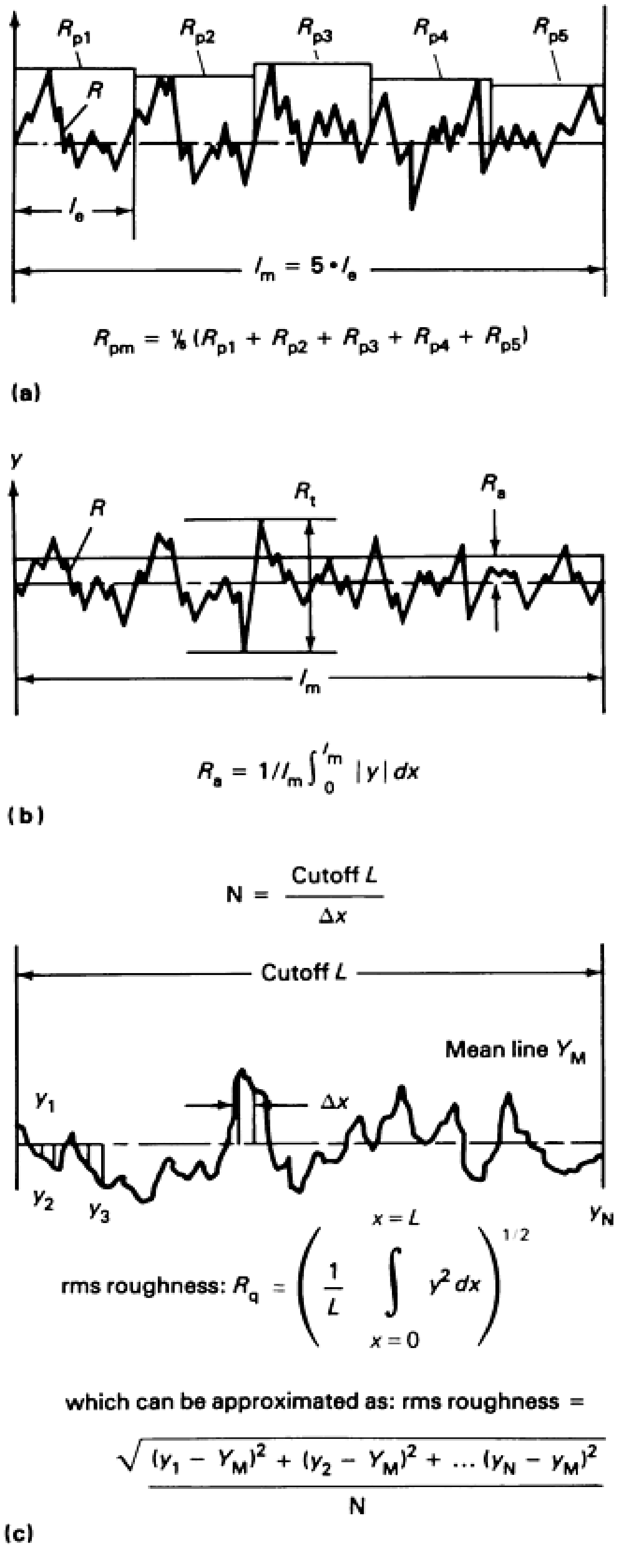


Fig. 5 Surface roughness parameters for bearings and bearing shafts. (a) Average leveling depth (R_{pm}), which is the mean of the individual leveling depths of five successive individual sample lengths (l_e) in the roughness profile. (b) Average roughness (R_a), which is the arithmetic mean of all deviation of the roughness profile *R*from

the mean line within the total measuring length (l_m). (c) Root-mean-square (rms) roughness (R_q)

Example 1: Maximum Roughness for Full-Film Lubrication (λ > 1) in the Case of a Steadily Loaded Journal Bearing with Good Alignment and Conformability.

Bearing dimensions include a shaft diameter, *d*, of 50 mm (2 in.); diametral clearance, c_d, of 0.050 mm (0.002 in.); and radial clearance, c_r, of 0.025 mm (0.001 in.). If (h_{min}/c_r = 0.1, then h_{min} is 0.0025 mm (0.0001 in.). For the λratio to be >1, the sum of the asperity heights must be <2.5 μm (<100 μin.). If the asperity heights on the shaft and the bearing are equal, the maximum asperity height on the shaft is <1.25 μm (<50 μin.). If the shaft is ground, it has a finish of R_{pm} < 1.25 μm (50 μin.), R_a < 0.3 μm (12 μin.), and σ_{rms} < 0.4 μm (16 μin.).

Significance of Low h_{min} Values. If the minimum oil film thickness drops to the extent that it is less than the value in Example 1, there is a progressive increase in asperity contact and the coefficient of friction will increase from ~0.001 to ≥0.05 (indicating that the heat generated during bearing operation has increased by a factor of 50).

Typical values of shaft surface finish are:

- R_a: 0.1 to 0.4 μm (4 to 16 μin.)
- R_q: 0.15 to 0.5 μm (6 to 20 μin.)
- R_{pm}: 0.4 to 0.6 μm (16 to 63 μin.)

As supplied, bearings can have a finish of 0.4 to 0.8 μm (16 to 32 μin.) R_a; however, because they are softer than the shaft, they rapidly polish down during break-in to have a roughness similar to that of the shaft.

Effect of Friction. When the mating surfaces in a steadily loaded journal bearing are fully separated by a film of oil, the friction is very low and depends primarily on the viscous shear of the oil film. The frictional heating can be calculated as a power loss that increases with speed and decreases with increased clearance, but that is substantially controlled by the value of h_{min}/c_r (Fig. 6). It should be noted that the chart in Fig. 6 assumes that the surfaces are sufficiently well aligned and smooth so that there is no asperity contact. The coefficient of friction is obtained from

$$\mu = \frac{H}{\pi d N W} \quad \textbf{(Eq 8)}$$

where *H* is the power loss (in watts), *d* is the diameter (in meters), *N* is the speed in (rev/s), and *W* is the load (in newtons).

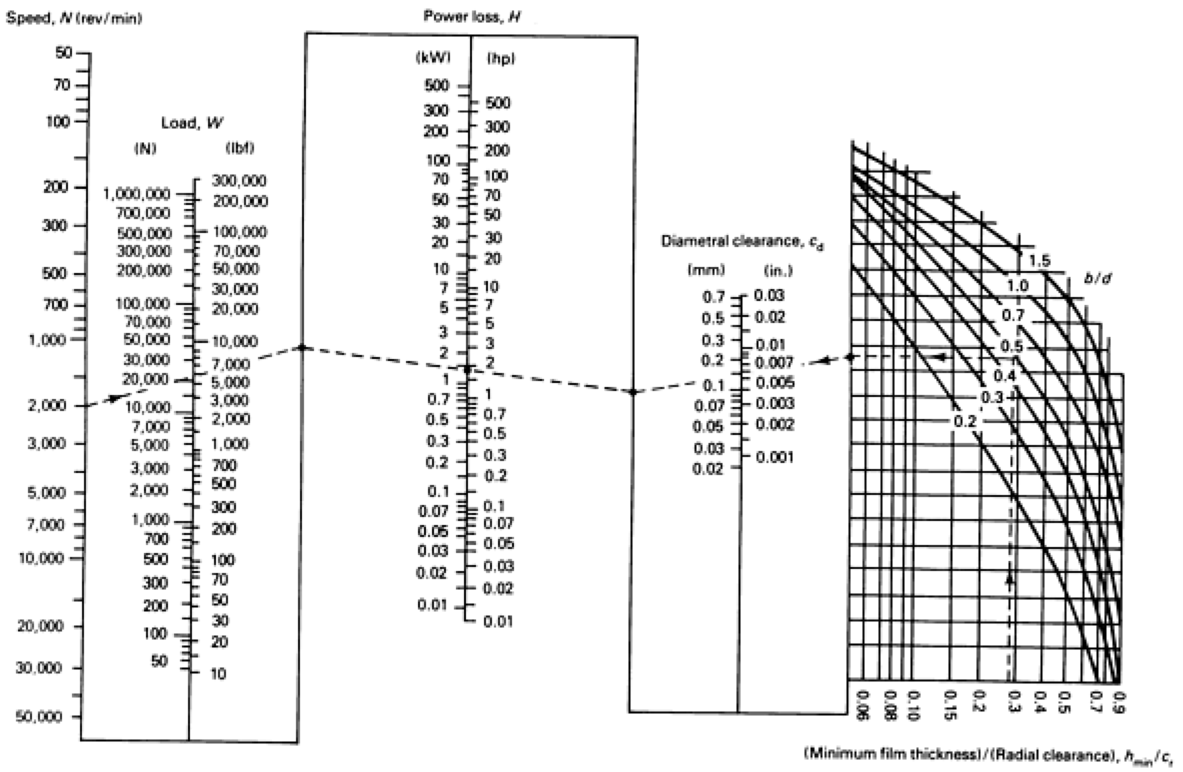


Fig. 6 Nomographs used to predict power loss in steadily loaded bearings. Source: Ref 4

Example 2: Calculation of Coefficient of Friction from Power Loss in a Full-Film Bearing.

The following parameters (Fig. 6) apply to an operational journal bearing:

Parameter	Value
Shaft diameter (*d*), mm (in.)	125 (5)
Diametrical clearance (C_d, mm (in.)	0.15 (0.006)
Speed (*N*), rev/s (rev/min)	33.33 (2000)
Load (*W*), N (lbf)	2.2×10^4 (5.0×10^3)
Width (*b*), mm (in.)	62.5 (2.5)
b/d	0.5
Specific load, MPa (psi)	2.8 (400)
Oil feed temperature, °C (°F)	50 (122)
Oil grade at 40 °C (105 °F), cSt	84
h_{min}/c_r	0.28
h_{min}, mm (in.)	0.021 (0.0008)
Power loss (*H*), kW (hp)	1.4 (1.9)

The coefficient of friction with hydrodynamic oil film is obtained by substituting the appropriate values into Eq 8:

$$\mu = \frac{1.4 \times 10^3\ W}{(3.1416)\ (0.125\ \text{m})\ (33.3\ \text{rev/s})\ (2.2 \times 10^4\ \text{N})} = 0.005$$

Friction and Wear under Mixed-Film Lubrication

The transition region between full-film lubrication (surfaces completely separated by fluid film) and boundary lubrication (load completely carried by rubbing contact) is known as mixed-film lubrication. Under these conditions part of the load is carried by asperity contact and part of the load is carried by the fluid film.

The coefficient of friction is often expressed in terms of a Stribeck curve plotted against the Sommerfeld number (Fig. 7). As the load increases, or speed and viscosity decrease, the coefficient of friction drops until a transition point is reached when the oil film is so thin that asperity interaction occurs. As the λR_{pm} ratio drops to <1, the asperities carry a progressively increasing proportion of the load and the friction rises to that of boundary lubrication (Ref 6). Because the rate of heat generation may increase 10 to 50 times, it is essential that for high-duty applications the proportion of time spent under boundary lubrication is minimized. The *PV* factor for that part of the load carried by the asperities must be within the capability of the bearing/shaft system to absorb/conduct away the heat produced without generating excessive temperatures.

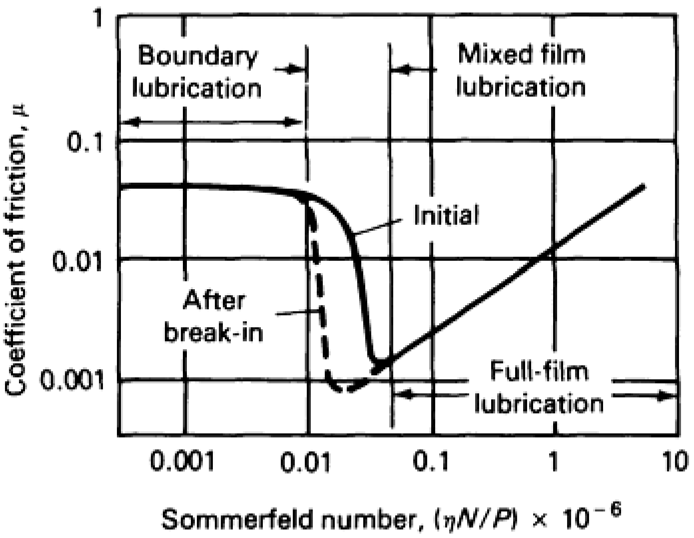


Fig. 7 Stribeck curve of coefficient of friction versus Sommerfeld number. (*S*), where *S* = η*N/P. N,* shaft speed; *P,* average pressure between shaft and bearing due to applied load; η, lubricant viscosity

For materials with good compatibility, the surfaces may break-in to the degree that asperity polishing reduces the effective height of the asperities. Thus, it may be found that the point of minimum friction for a well run-in bearing may occur at one-half or one-third of the Sommerfeld number for new components. Hence, for applications where frequent starts and stops under load are required, babbitt metal is a preferred material.

If, however, start-up specific loads are >1.5 MPa (>200 psi), significant wear may occur during start-up and wiping may occur on extended rundown (particularly if the shaft is hot; that is, 120 °C, or 250 °F). In these circumstances it is advisable to inject high-pressure oil ("jacking oil") for start-up and rundown. For continuous high steady loads (that is, 10 to 20 MPa, or 1.5 to 3.0 ksi) and low speeds (such that $h_{\mathrm{min}}/c_{\mathrm{r}}$ is unacceptably low), it may be necessary to inject high-pressure oil continuously (that is, a hydrostatic bearing configuration); an example is a rolling mill bearing. The oil supply pressurefor a hydrostatic bearing must be two to three times the maximum specific load. The friction and wear for a hydrostatic bearing or for bearing with an effective oil jacking oil system will be very low (that is, μ< 0.001 at low speeds).

Oil Additives and Surface Treatments

Although reference has been to "asperity contact," there will be lubricant molecules attaching to the asperities that will, to a greater or lesser extent (depending upon the formulation of the oil), modify the friction an wear. Thus, it is difficult to give even general guidelines concerning wear rates under boundary or mixed-film conditions because these area critically

dependent, not only upon the bearing material and the mating surface but also upon the lubricant itself and the extreme-pressure additives present in the lubricants formulated for any specific application. Wear cn also be significantly modified by surface treatment of the mating surface (for example, hardening, nitriding, an so on).

Dynamic Loads

In reciprocating machinery (internal combustion engines, reciprocating compressors, and so on), the loads on the bearings vary in both magnitude an direction during each revolutions (two-stroke engines) or alternate revolutions (four-stroke engines) (Fig. 8). Details of a procedure for calculating the performance of a dynamically loaded bearing can be found in Ref 7.

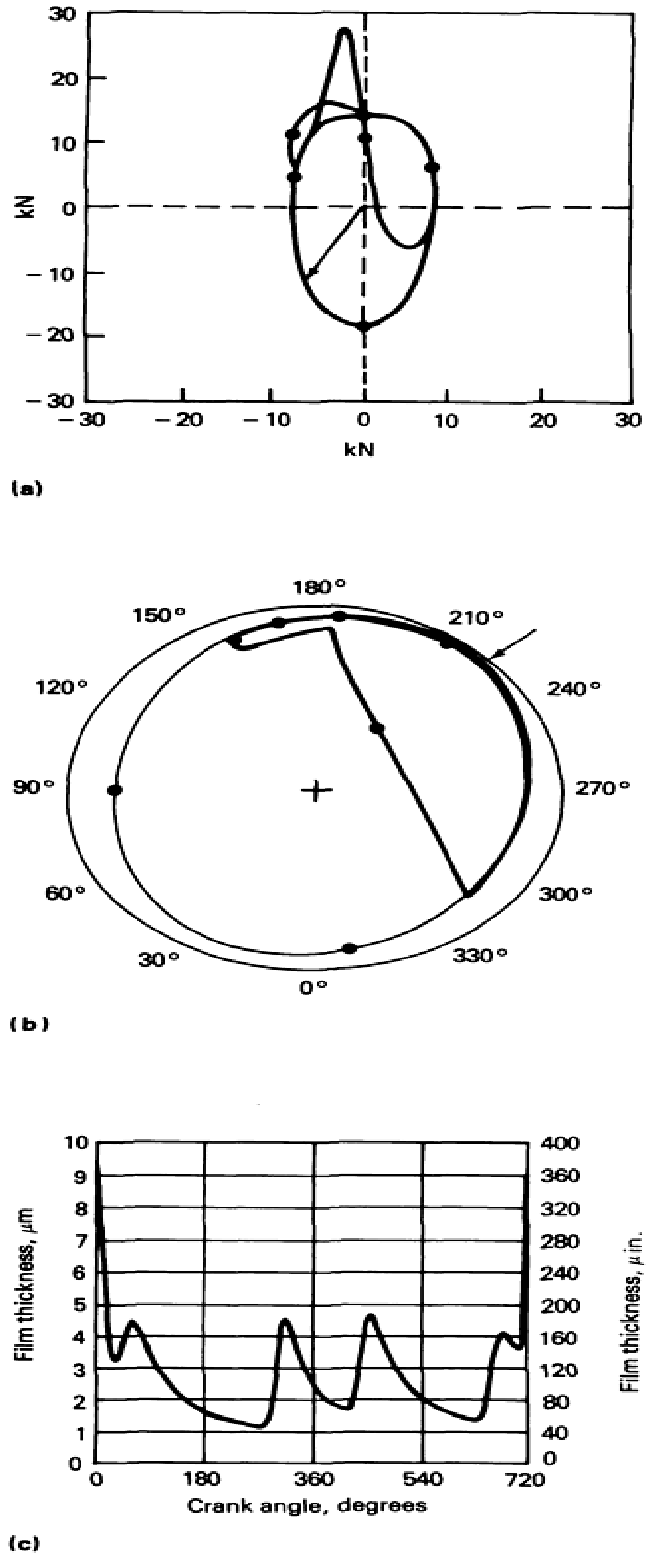


Fig. 8 Monitoring of parameters that measure the performance of a dynamically loaded bearing. (a) Polar load diagram showing how load on the connecting rod bearing varies in magnitude and direction during consecutive revolutions (720°) of the crankshaft in a four-stroke gasoline engine operating at 5250 rev/min. (b) Journal center orbit within the bearing clearance circle. (c) Oil film thickness. The minimum oil film thickness drops to <1.5 μm (<60 μin.) for two periods of a few milliseconds duration during each cycle. Source: Ref 7

Rapid displacement of the journal within the bearing is resisted by the lubricant, and relatively high loads can be carried for brief periods during each revolution. Typically, the maximum specific loads can each 30 to 60 MPa (4.5 to 9 ksi) in

connecting rod bearings, and thus the material must be chosen to have sufficient fatigue strength. Additional information is available in the article "Friction and Wear of Sliding Bearing Materials" in this Volume.

As is evident from their reliability, bearings for reciprocating machinery operate for the majority of their life under full-film lubrication conditions. However, the development of the internal combustion engine and, in particular, the requirement for good fuel economy has resulted in a reduction in the size and the weight of engines coupled with an increase in specific rating. This has resulted in bearings being required to operate with progressively thinner minimum oil film thickness. For the majority of each revolution, the surfaces are separated. However, intermittent asperity contact may occur during various periods of the operational envelope:

- Initial break-in
- At start-up
- At sustained high speed

Sustained High-Speed Operation. The minimum oil film thickness generally occurs at high speeds due to the inertia forces being at maximum. With time, the temperature rises (over a few minutes) to the point that the correspondingly reduced viscosity may be insufficient to maintain an adequate oil film, and asperity contact and wear may occur.

Start-Up Operations. During start-up, there may not be adequate oil supply because in many machines the pump is driven by the shaft; thus, for the first few revolutions, the bearing may have to operate with the oil that remains by capillary action from previous operation. If the engine is cool, there is generally sufficient lubrication. However, repeated "hot" starts may result in excessive wear. At very low temperatures (-20 °C, or -4 °F), the oil may become waxy and the oil pump may not be able to circulate the oil, necessitating the use of a special winter grade oil.

Initial Break-In Operation. The initial break-in period is an important factor in bearing performance. Minimum oil film thickness in automotive internal combustion engines may be as low as 0.001 mm (0.00004 in.), and it is important that crankshafts are not only ground, but also lapped in order to minimize the height of the surface asperities above the mean level of the surface. For such a thin oil film, the typical height of the asperities on the crankshaft and bearing should be <0.5 μm (<20 μin.) R_{pm} after break-in (that is, the sum of e_s + e_b should be <1 μm, or 40 μin.). It may not be economical to achieve such fine finishes in production, and it is normal to rely on the compatibility of bearing material and shaft to achieve a mutual polishing to reduce the height of the asperities and produce a "plateau" or "negatively skewed" distribution of surface heights (Ref 8, 9).

Effect of Operating Conditions on Lubricated Bearing Performance

Imperfections in Geometry. Any deviation from plane, cylindrical, and parallel surfaces in excess of the minimum oil film thickness will result in asperity contact. If initial operating conditions are benign, opportunity exist for asperity polishing (break-in compatibility) or local wear (bedding-in comformability). The bearing assembly can accommodate minor misalignments and faults in assembly (for example, bore distortion, waviness, "barrel"-shaped shaft, and so on). If the defects are excessive, too much of the load may be carried by localized rubbing contact, which leads to overheating, more rapid wear, scuffing, or possibly seizure. Some bearing materials (for example, babbitt metal) may accommodate relatively large geometrical defects by "wiping"(partial melting) or plastic deformation.

Dirty Environment. Any hard particulate contaminants with dimensions in excess of the minimum oil film thickness can be expected to cause plowing wear or scoring of the soft surface of the bearing. Larger particles may become embedded in the bearing and lap or score the mating shaft. Bearing assemblies should be sealed against the ingress of particulate contaminants; if the assembly is lubricated, the oil should be filtered.

Corrosion. Gases and fluids that contaminate lubricants may reduce bearing life by corrosion. Typical examples include water in turbines causing tin oxide formation on tin-base babbitt, and refrigerants corroding babbitt containing copper. Lubricants for engines are provided with additives to inhibit oil oxidation and to minimize corrosion of trimetal bearings; however, dilution with fuel, water, or excessive piston blow-by gas may neutralize the effect of the additives.

Dynamic Loading. Repetitive loading may lead to fatigue of the softer bearing materials. This may occur in reciprocating machinery or occasionally in a high-speed rotating plant due to lack of balance or oil film instability (whirl).

Lack of balance is observed at synchronous frequency and may be attributed, for example, to a thermally bent shaft (labyrinth seal rub) or unsymmetrical aggregation of deposits, erosion on turbine blades, and so on.

Oil film whirl (or "whip") generally occurs at one half synchronous frequency or less and requires a redesign of the bearing (that is, multi-lobe, tilting pad journal, and so on).

Cavitation Erosion. Shaft vibration resulting from dynamic loading may contribute to cavitation erosion of soft bearing surfaces (for example, trimetal bearings), particularly if clearance is excessive (>0.002 mm/mm, or 0.002 in./in.). However, cavitation erosion is more often due to interruptions in the oil supply (as in partially grooved crankshaft bearings) or low oil supply pressure.

Inadequate Lubrication. If wear and friction are to be reduced to minimum, it is essential that sufficient lubricant be supplied to separate the asperities of the mating surfaces. In some machinery, experience has shown that it is possible to obtain an adequate bearing life even when the bearing is operated under oil starvation conditions. Traditional reciprocating steam engines, for example, were able to operate with oil supplied from wick or drip feed lubricators that were topped up hourly or daily. The *PV* factor for such applications was generally <1 MPa · m/s (<3.5×10^4 psi · ft/min), and provision was made to adjust the clearance (take up the slack). For bearings required to operate at a much higher *PV* rating, the oil should be supplied continuously at a minimum pressure of 105 kPa (15 psi) at the entrance to the bearing. When allowance is made for pressure loss in oilways, such conditions generally require a minimum of 210 kPa (30 psi) pressure leaving the oil pump for rotating plant and 310 to 415 kPa (45 to 60 psi) for reciprocating engines. A temporary loss of pressure, even for a few seconds, may lead to overheating and seizure in highly rated bearings.

For light duties, where the *PV* factor is low and heat can be dissipated by conduction, lubricant is provided principally to reduce friction. In such cases, it may be in the form of splash feed, drip lubrication, wick, or self-lubricated (for example, oil-impregnated porous bronze) system. For high loads, particularly those in oscillating pivots, grease is preferred. The presence of indentations or deep grooves to retain and distribute grease may extend the period of bearing operation before regreasing is required.

All bearings in which the asperities are not completely separated by a lubricant film can be expected to wear at a rate that will depend on the application, materials, temperature, lubrication, and so on. Because of the wide range of operating environments, the data given in this article should be considered as an initial guide only. The suitability of a bearing system for any particular application must be based on relevant experienced or established by test.

References

1. "Design and Material Selection for Dry Rubbing Bearings," Data item 87007, ESDU International Ltd.
2. K.C. Ludema, Failures of Sliding Bearings, *Failure Analysis and Prevention,* Vol 11, 9th ed., *Metals Handbook,* American Society for Metals, 1986, p 484
3. "Calculation Methods for Steadily Loaded Pressure Fed Hydrodynamic Journal Bearings," Data item 84031, ESDU International Ltd.
4. F.A. Martin and D.R. Garner, "Plain Journal Bearings under Steady Loads: Design Guidance for Safe Operation," Paper C313/73, lst European Tribology Congress 1973, Institute of Mechanical Engineers, 1974
5. "1978 Surface Texture," B46.1, American National Standard Institute, paragraph C.3.2.8, p 28
6. H. Christensen, Some Aspects of the Functional Influence of Surface Roughness in Lubrication, *Wear,* Vol 17, 1971, p 149-162
7. J.M. Ross and R.R. Slaymaker, "Journal Center Orbits in Piston Engine Bearings," 690114, Society of Automotive Engineers
8. "1978 Surface Texture," B46.1, American National Standards Institute, paragraph C.3.3.4 and Fig C14, p 32
9. D.R. Bury, D.R. Eastham, and J.M Conway-Jones, Relocatable 3-D Mapping of Tribological Surfaces, *J. Phys. D, Appl. Phys.,* Vol 25, 1992, p A297-A284

Friction and Wear of Gas-Lubricated Bearings

Dudley D. Fuller, Columbia University

Introduction

AIR-LUBRICATED BEARINGS have along been a source of interest, but it was not until World War II that a serious concerted effort was initiated to develop the potential of gas-lubricated bearings. The occasion was the Manhattan Project devoted to the creation of the atomic bomb. It is well known that for both domestic power generation and for military purposes, a form of enriched uranium hexafluoride is required (Ref 1). The most significant source of this enriched fuel is the gaseous diffusion plant. These plants provide the enriching process by pumping uranium hexafluoride through a cascade of successive diffusion stages. Several thousand in series may be needed to achieve the desired composition. To avoid unacceptable high energy losses and inevitable contamination of the product associated with the use of conventional oil-lubricated bearings and seals, the pumping equipment bearings are instead lubricated by gaseous uranium hexafluoride.

In 1957 a coordinated research program on gas lubrication was initiated by the Office of Naval Research (ONR). The activity soon became international--with an equivalent group in the United Kingdom working, sharing, and meeting with ONR alternately in England and the United States. Two international symposia were held (Ref 2, 3) with participants from all over the world. In 1981, the sponsorship of the coordinating meetings was assumed by Wright-Patterson Air Force Base, where it continues as of this writing.

Presently, gas-lubricated bearings are found in:

- High-speed grinding spindles
- Sliding ways on machine tools for avoiding stick-slip vibration
- Bearings for precise linear and rotational indexing metrology devices
- Hermetically sealed high-speed blowers and compressors using the process gas as the lubricant, essential for gas-cooled reactors and for cryogenic and pyrogenic turbomachinery where great temperature extremes are encountered
- High-precision inertial guidance instruments such as gyroscopes and accelerometers
- Foil-type gas bearings for air-cycle turbomachines on aircraft and space-borne power-generating components (the immediate goal is to produce 300 kW, or 400 hp, from a unit with a volume of only 0.03 m^3, or 1 ft^3)
- Computer peripheral devices, including magnetic memory devices, tapes, disks and drums, read-write heads "flying" with an air gap <0.25 μm
- High-speed dental and orthopedic drills and cutters that operate at up to 500,000 rpm with quiet, oil-free performance

At the other end of the precision spectrum, externally pressurized air-lubricated membrane bearings are used commercially to support and position large machine tools, heavy die blocks, and even full-scale railroad cars, on the factory or shop floor.

Advantages and Disadvantages

The advantages of gas-lubricated bearings over liquid-lubricated fluid-film bearings are now well understood. These include:

- Cleanliness. Elimination of contamination caused by more traditional lubricants

- Reduction (often elimination) of the need for bearing seals
- Stability of lubricant. No vaporization, cavitation, solidification, decomposition, or phase change over extreme ranges of temperature, from cryogenic (-270 °C, or -450 °F) up to approximately 1650 °C (3000 °F). Operation at these extremes of temperature is a current research goal
- Low friction and heating with little or no cooling generally required. Permits practical attainment of high speeds (700,000 rpm)

Disadvantages of gas-lubricated bearings are recognized as resulting from the relatively low viscosity and damping of gas films. Thus, gas-lubricated bearings have a reduced load-carrying capacity compared to liquid-lubricated bearings, especially with self-acting or hydrodynamic bearings. For acceptable application, therefore, the bearings are necessarily larger and operate with thinner hydrodynamic films than their liquid-lubricated counterparts.

Thinner films demand closer control of manufacturing tolerances, surface finishes, and possible thermal and elastic distortions and alignments, to prevent rubbing contact. With compliant surface bearings, as with foil bearings and membrane bearings, rigid specifications regarding design and manufacture can be dramatically relaxed. The membrane bearing, for example, operates very satisfactorily over a typical factory floor.

The low damping of the gas film makes it necessary to carefully analyze the dynamic characteristics of the mechanical system employing the gas bearing, since if a critical speed or instability is encountered, there may not be enough damping to suppress it or control it. With liquid-lubricated bearings these instabilities might not have been suppressed or passed over unnoticed because of the greater damping action that inherently exists with liquids. Much recent research has been devoted to the dynamics of gas bearings and their associated mechanical systems.

Gas-lubricated bearings have been characterized as being less forgiving than oil-lubricated bearings. This is certainly true for self-acting or hydrodynamic bearings. They are less forgiving of errors in estimating loads, of deviations from specifications during manufacture and installation, and of distortions and dirt that may afflict the rotor.

Compressibility Numbers

Gas is, of course, compressible and this effect must be included in the derivations for various forms of bearings--whether hydrodynamic (self-acting) or hydrostatic (externally pressurized). The extent of compressibility (represented by the compressibility number, Λ) is determined by a dimensionless group of parameters that actually evolves from the mathematical analysis. It takes on several forms depending on the geometry of the bearing.

For journal bearings (Fig. 1), the value of $\Lambda$is:

$$\Lambda = \frac{6\mu\omega}{p_a}\left(\frac{r}{c}\right)^2 \qquad \textbf{(Eq 1)}$$

where c is the machined-in radial clearance, p_a is the absolute ambient pressure, $\omega$is the angular velocity of the journal (rad/s), $\mu$is the absolute viscosity, and r is the radius of the shaft.

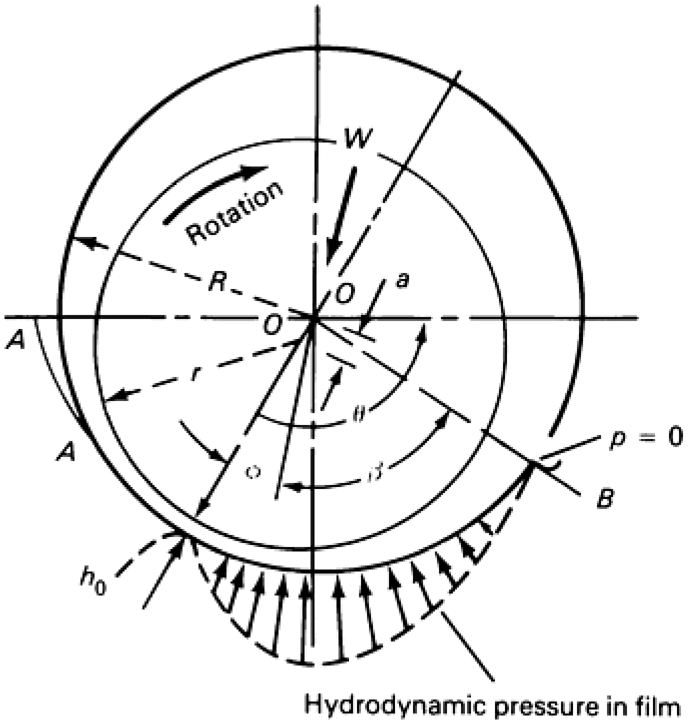


Fig. 1 Journal in full 360° bearing

For a tilting-pad thrust bearing (Fig. 2), the value of Λ is:

$$\Lambda = \frac{6\mu u l}{h_2^2 p_a} \quad \textbf{(Eq 2)}$$

where l is the length of the pad in the direction of sliding (sometimes designated as L), and h_2 is the minimum film thickness (sometimes referred to as h_0), and u is the velocity of the runner past the shoe.

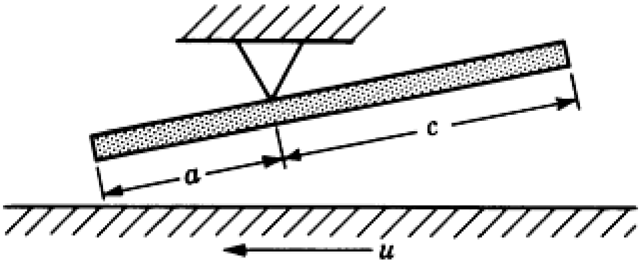


Fig. 2 Representation of pivoted shoe

For a Rayleigh step bearing (Fig. 3), the value of Λ is:

$$\Lambda = \frac{6\mu U l}{h_2^2 p_a} \quad \textbf{(Eq 3)}$$

where l is the length of the pad in the direction of sliding, U is the velocity of the runner past the pad, and h_2 is the minimum film thickness.

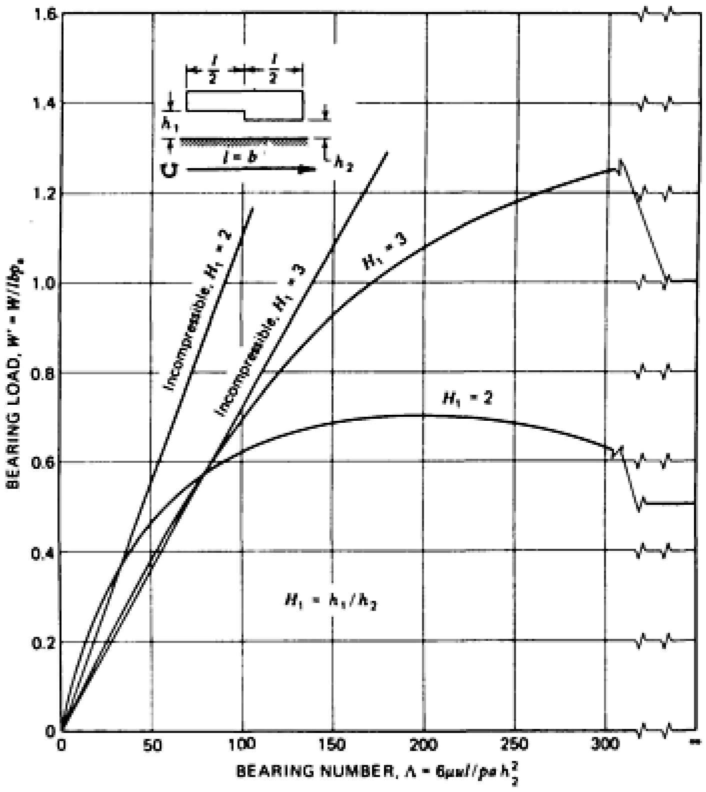


Fig. 3 Effect of bearing number on isothermal load for square-step slider bearings with film thickness ratios h_1/h_2 = 2 and 3. Source: Ref 4

When Λ approaches 0, operation of the gas-lubricated bearing approaches that of the liquid-lubricated (incompressible) case. As a Λ gets larger, as with lower ambient pressure or higher speed, the compressibility effects become very significant and must be included. For example, with journal bearings it can be shown that the equations for determining the load-carrying capacity for liquids and gases are essentially the same up to a value of Λ= 1 (Fig. 4). For tilting-pad bearings, the identity remain up to a value of about 15 (Fig. 5).

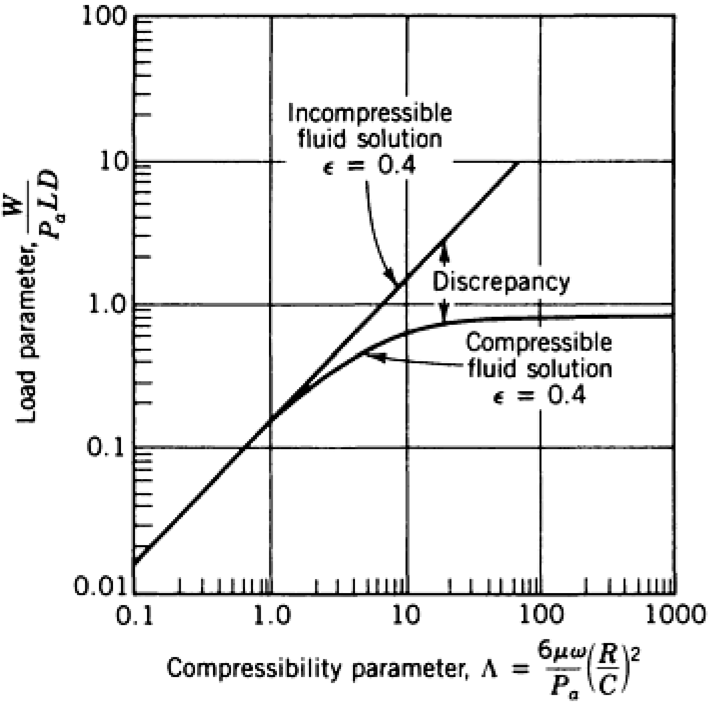


Fig. 4 Discrepancy in load capacity between results based on incompressible and compressible lubricants. Source: Ref 5

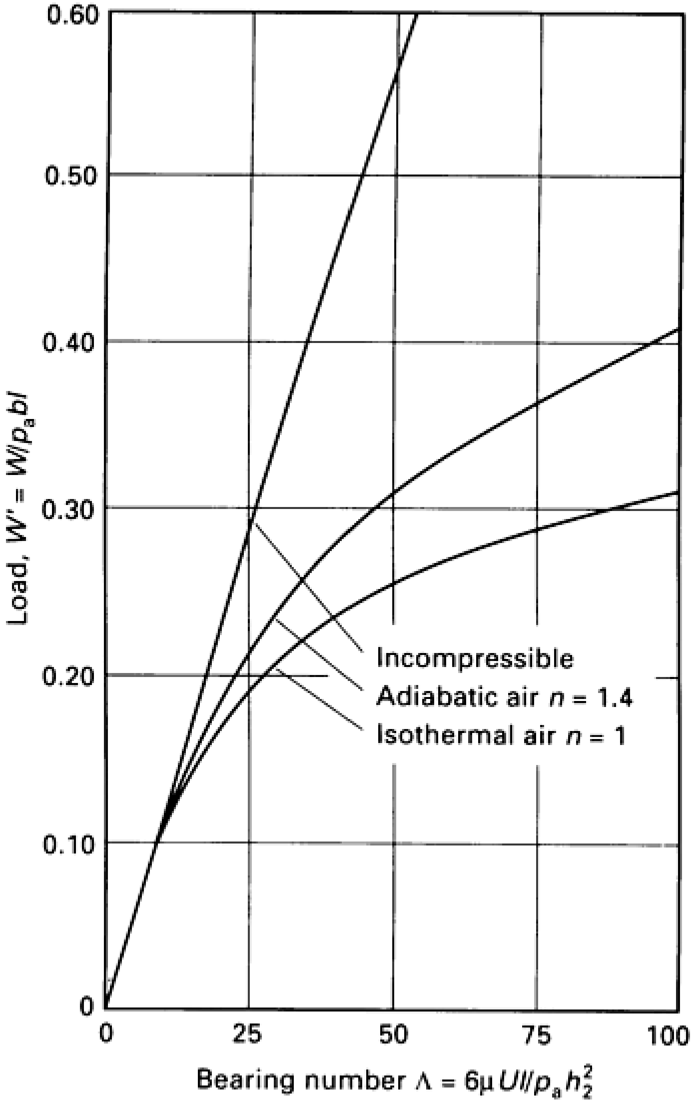


Fig. 5 Effect of bearing number on isothermal and adiabatic load for plane slider bearing operating in air with film thickness ratio h_1/h_2 = 2. Length-to-breadth ratio l/b = 1. Source: Ref 6

An interesting experimental demonstration is illustrated in Fig. 6, which shows the relationship between load-carrying capacity, *W*, and Λ for a hydrodynamic gas-lubricated journal bearing. Speed, viscosity, and eccentricity ratio were held constant while the ambient pressure was reduced, thus producing higher values of Λ. Notice that when the ambient pressure was reduced to 9% of atmospheric, the load-carrying capacity was lowered by 40%.

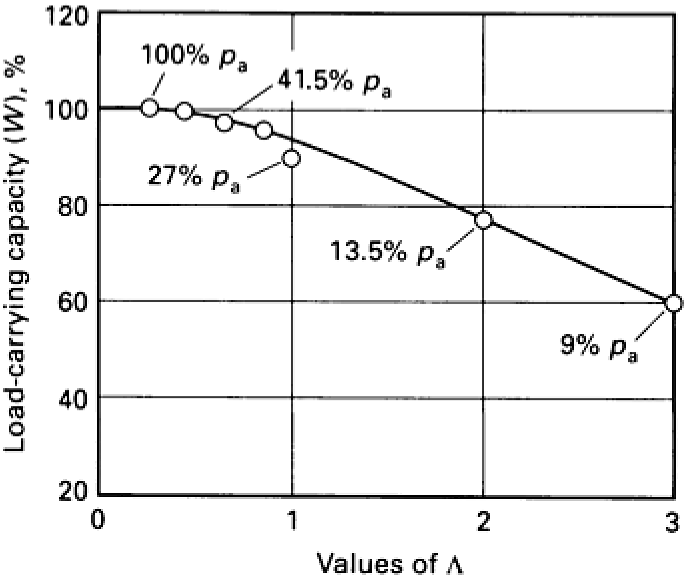


Fig. 6 Example of relationship between load-carrying capacity, *W*, and Λ for self-acting gas-lubricated journal bearing. Speed, viscosity, and eccentricity ratio held constant. Ambient pressure varied. Source: Ref 7

Definitions of Eccentricity Ratio and Clearance Modulus

In Fig. 1, *O* represents the center of the bearing and *O*' the center of the shaft. The distance *O*-*O*' is also designated as *a;* for light loads and high operating speeds, the center of the shaft and bearing coincide, and *a* approaches zero. The ratio of the distance *a* for any given operating condition to the machined-in radial clearance *c* of the bearing is called the eccentricity ratio, ϵ.

For heavy loads or extremes of operation, such as low speed or very low viscosity, the journal becomes more eccentric in the bearing and the distance *a* increases. The limit is where the journal just begins to make solid contact with the bearing, or the distance *a* equals the radial clearance. The eccentricity ratio is then 1. The radial clearance in the journal bearing is often designated as *mr,* where *m* is the clearance modulus of the bearing and *r* is the radius of the shaft or journal. The value of *m* will usually range between 0.0005 and 0.003 mm/mm, with typical industrial-type bearings running between 0.001 and 0.0025 mm/mm.

In the analysis of fluid-film bearings using incompressible lubricants, the continuity relationship was satisfied by saying that the volume of lubricant leaving the bearing was the same as the volume entering. With gas, of course, the volume changes, so that continuity must be based on equality of mass. Thus the mass of lubricant leaving equals the mass entering.

The basic Reynolds equation then becomes (Ref 8):

$$\frac{\partial}{\partial x}\left[\frac{\rho h^3}{\mu}\frac{\partial p}{\partial x}\right] + \frac{\partial}{\partial z}\left[\frac{\rho h^3}{\mu}\frac{\partial p}{\partial z}\right] = 6U\frac{\partial \rho h}{\partial x} \quad \textbf{(Eq 4)}$$

where ρ is the mass density, *x* is the coordinate along the film in the direction of motion, *z* is the coordinate along the bearing dimension perpendicular to the direction of motion (sometimes called the length of the bearing, *l,* or the width of the bearing, *b*), *p* is the differential pressure, and *h* is the local differential film thickness.

Except for a few restricted solutions with no side flow in the *z* direction, solutions are only possible through extended iterative computer solutions. The order of magnitude of the solutions, however, can often be checked by reference to results for incompressible lubricants, especially for cases of low Λ(Ref 9).

Tilting-Pad Bearings

Some marked differences in behavior exist between tilting-pad gas-lubricated bearings (Fig. 2) and liquid-lubricated bearings. The expected pressure is somewhat modified, as shown comparatively in Fig. 7. The peak pressure is shifted toward the trailing edge of the shoe. This means that the pivot position for a gas-lubricated bearing is not the same as for a liquid-lubricated bearing. Figure 8 shows quantitative values for a tilting-pad gas-lubricated bearing for various values of Λ(Eq 2). In this case, $\Lambda$was reduced by lowering the ambient pressure.

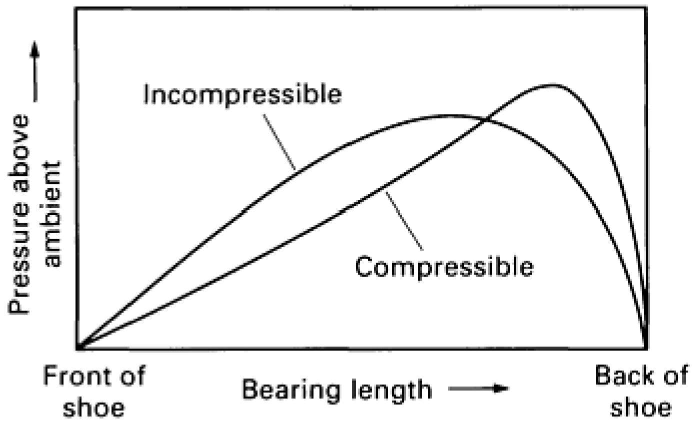


Fig. 7 Comparative pressure distribution for compressible and incompressible lubrication. Source: Ref 6

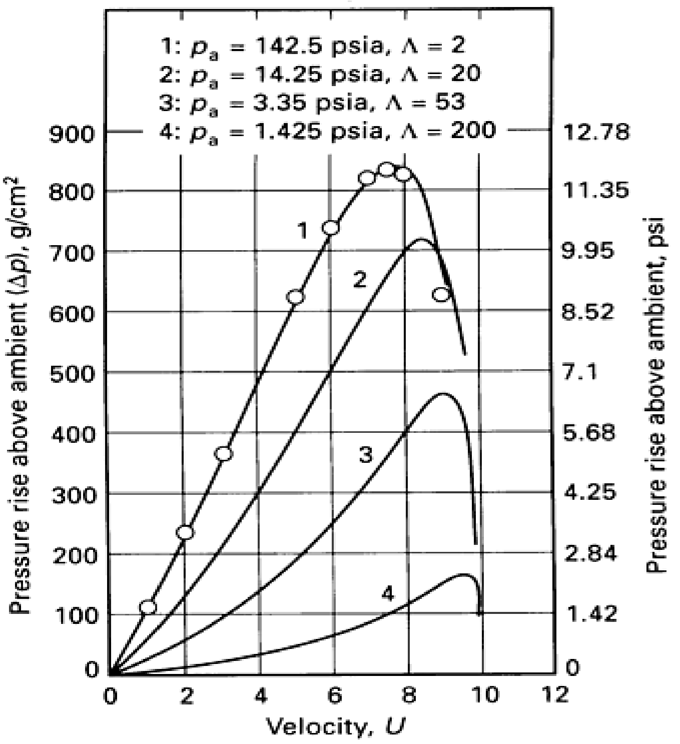


Fig. 8 Gas pressure distribution in flat tilting-pad bearing. Note open circles (data points) for theoretical liquid case. Source: Ref 7

Notice that for high values of Λ, the center of pressure moves toward the trailing edge or the back of the pad. For each designated operating condition, the pivot should be located at the center of pressure. Thus, if the ambient pressure, the speed, or the load is subject to change, a fixed pivot position will be optimum for only one set of operating conditions. This is a disadvantage.

In contrast, with incompressible lubricants, the tilting pad will react to a change in load or speed so that the ratio of inlet-to-outlet film thickness remains constant, even with the minimum film thickness getting smaller with increasing load. Thus, with liquid lubricants, once the pivot position is chosen, it will remain optimized for all changes of load, speed, or viscosity.

With gas-lubricated tilting-pad bearings, yet another complication arises. As the load increases, the minimum film thickness decreases, but since the pivot location is fixed, the pad angle of inclination begins to get smaller and eventually becomes parallel to the runner. An unstable condition will result, and the bearing will be unable to support the load and will collapse completely.

Collapse of the gas film will not occur for a convex curved surface on a tilting pad (Fig. 9). Because of the geometry, neglecting the influence of surface roughness asperities, a wedge must always be developed no matter how thin the film. An experimental study by Brunner *et al.* (Ref 12) shows that for a pad 19 mm ($\frac{3}{4}$ in.) wide by about 13 mm ($\frac{1}{2}$ in.) long in the direction of motion, even a crown height, δ, of as little as 0.25 μm will prevent collapse of the film and produce a load-carrying characteristic that continues to rise as the film reduces. Experience has shown that crown heights for optimum load-carrying capacity are approximately equal to the minimum film thickness, or about 7.5 to 10 μm.

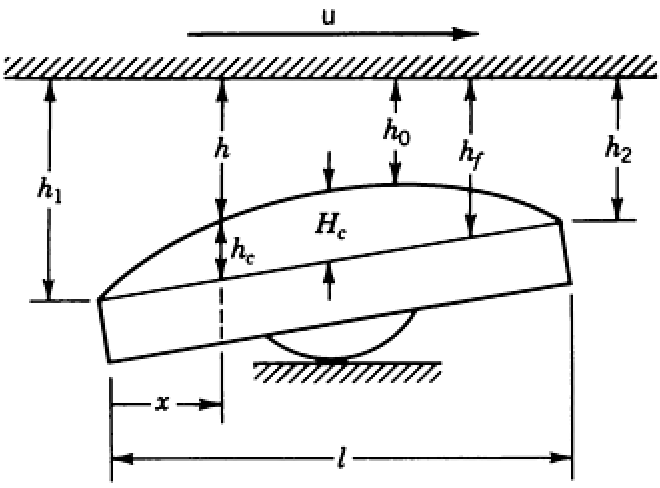


Fig. 9 Nomenclature for crowned tilting-pad bearing. Source: Ref 11

A thorough analysis of optimum crown heights and pivot positions can be found in Ref 4. Experimental verification of the theory has been especially good. Figure 10 shows values of the dimensionless load-carrying function as a function of the compressibility number for shoes having cylindrical crowns. The pivot position is not identified, but a value of about 0.6 is typical. That means, in Fig. 2, *c* is 0.6 of the length in the direction of motion and *a* is 0.4 of the length. The pivot is always between the midpoint and the trailing edge.

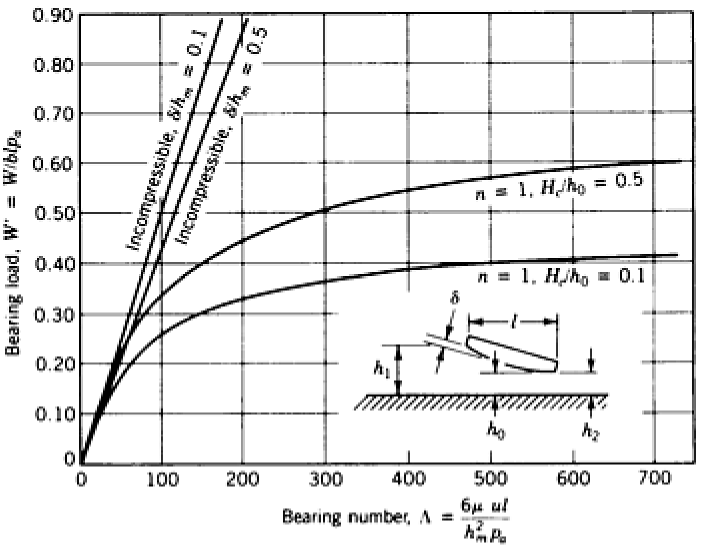


Fig. 10 Effect of bearing number on isothermal load for cylindrically crowned slider bearing, b/l = 1, h_1/h_m = 2, and H_c/h_0 = 0.1 and 0.5. Source: Ref 4

Effect of Mean Free Path

When gas films become very thin, as for example, when operating in a partial vacuum, the gas may no longer be considered as a continuous fluid, having the bulk viscosity of that fluid. Effects take place that can no longer be explained by continuous theory. Slip occurs at the boundary between the bearing surface and the gas and must be introduced into the calculation.

The measure of this slip condition is determined by the Knudsen number (K_n). This is the ratio of the mean free path of the gas (λ) to the film thickness in the bearing (h). Thus, $K_n = \lambda/h$. When K_n is less than 0.01, the gas flow can be treated as a continuum and slip flow ignored. When K_n is in the range of 0.01 to 1.5, slip flow increases and should be considered. When K_n is greater than 1.5, slip flow becomes significant and fully developed molecular flow results.

Typical values of mean free path at atmospheric conditions are:

Gas	Typical mean free path value	
	in. × 10^{-6}	**μm**
Hydrogen	4.43	0.1125
Helium	7.32	0.186
Air	2.52	0.064
Neon	5.20	0.132

Thus, with an air-film thickness of 2.5 μm (0.0001 in.), the value of K_n would be 2.52 × 10^{-6}/100 × 10^{-6}, or 0.025. The slip effect would be small but not negligible.

Many tilting-pad-type thrust bearings with air lubrication are used in magnetic storage devices for computers, where they "fly" at film thicknesses of about 0.5 μm (20 μin.). At these very small film thicknesses, surface roughness can become a significant factor because no surfaces are perfectly smooth. This effect can be analyzed and included in predicting load-carrying capacity.

Rayleigh Step Bearings

This type of bearing is shown in Fig. 3. Analytical results for a square-step bearing (width equals length) are shown in Fig. 11 for lubrication with gas. A dimensionless load function *W*' is plotted against the compressibility number established for step bearings (see Eq 3).

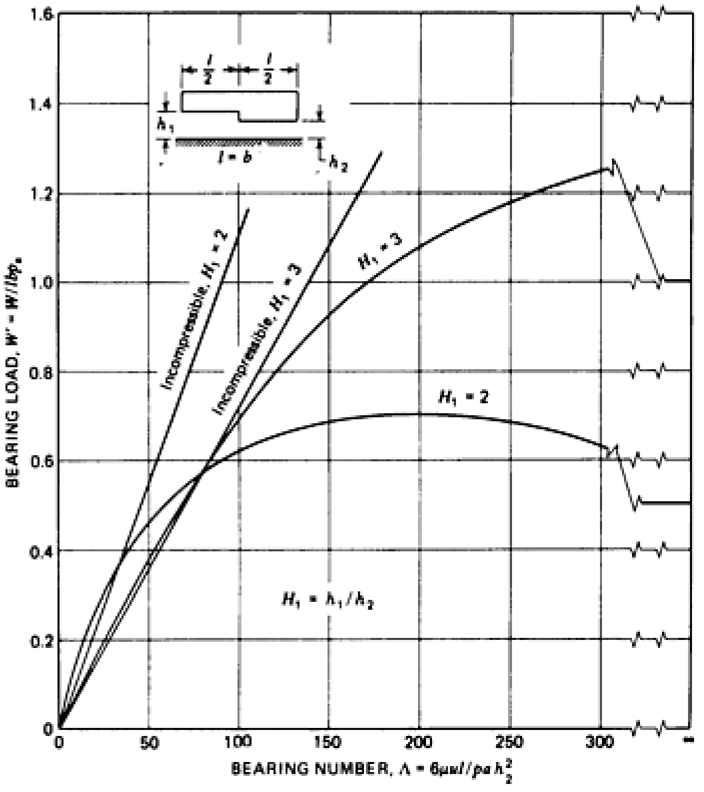


Fig. 11 Effect of bearing number on isothermal load for square-step slider bearings with film thickness ratios h_1/h_2 = 2 and 3. Source: Ref 4

Journal Bearings

This was the geometry first used to demonstrate true load-bearing ability by Kingsbury in 1897 (Ref 13). The same concept was used by Brubach (Ref 14) in utilizing inexpensive hypodermic syringes as air bearings. These are precision-made plungers in closely fitting barrels that perform remarkably well as hydrodynamic journal bearings.

Brubach has used these syringes as bearings for small centrifugal blowers running at 6000 rpm (Fig. 12). Perhaps his most important application has been in the measurement of gas leaks at very low pressures. Figure 13 shows the apparatus used for this purpose. The plunger of the syringe is caused to rotate by air impinging on the Scotch tape vanes. Due to this rotation, the plunger acts as an air bearing and is separated from the barrel by a hydrodynamic air film. Because the surfaces are separated from the barrel by a thin air film, the force required to move the plunger along the axis of the barrel actually approaches zero.

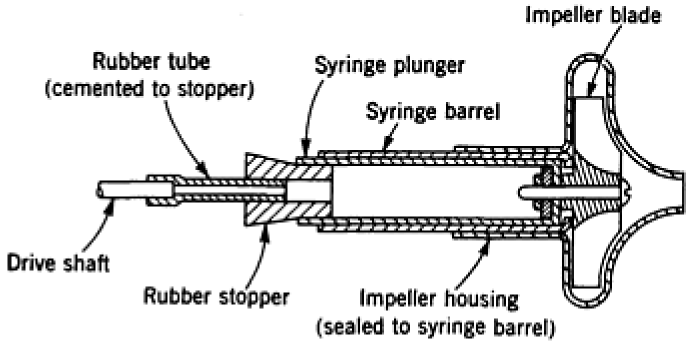


Fig. 12 Centrifugal blower using hypodermic syringe as air-lubricant bearing. Source: Ref 14

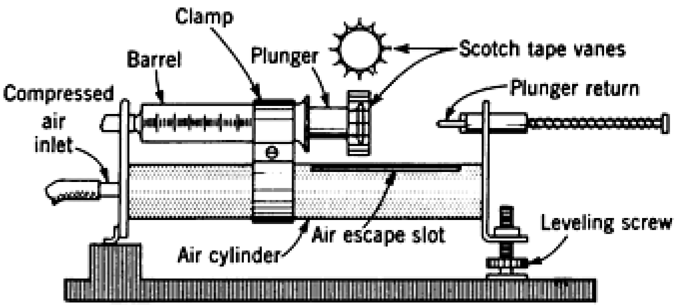


Fig. 13 Sensitive pressure-measuring apparatus using hypodermic syringe as air-lubricated bearing. Source: Ref 14

As an illustration of the sensitivity of this apparatus, it was connected to a tank of 100 L (26 gal) capacity. Then 0.1 mL (0.06 in.^{3}) of air was introduced into the tank. Immediately the plunger moved a corresponding 0.1 mL (0.06 in.^{3}). The theoretical pressure increase resulting from the injection of 0.1 mL (0.06 in.^{3}) into 100 L (26 gal) is about 0.01 mm (0.0004 in.) of water, and yet the low friction of the rotating plunger permitted it to respond instantly to this exceedingly small pressure differential.

Journal bearings are also influenced by compressibility of the gaseous lubricant. The value of Λ was given in Eq 1. Ford, Harris, and Pantall (Ref 7) show experimental correlation of the load-carrying parameter *W'* plotted against Λ (Fig. 14). For values of Λ greater than 1, the relationship is no longer linear. The deviation from a straight line indicates the point at which the compressibility effects make their influence felt.

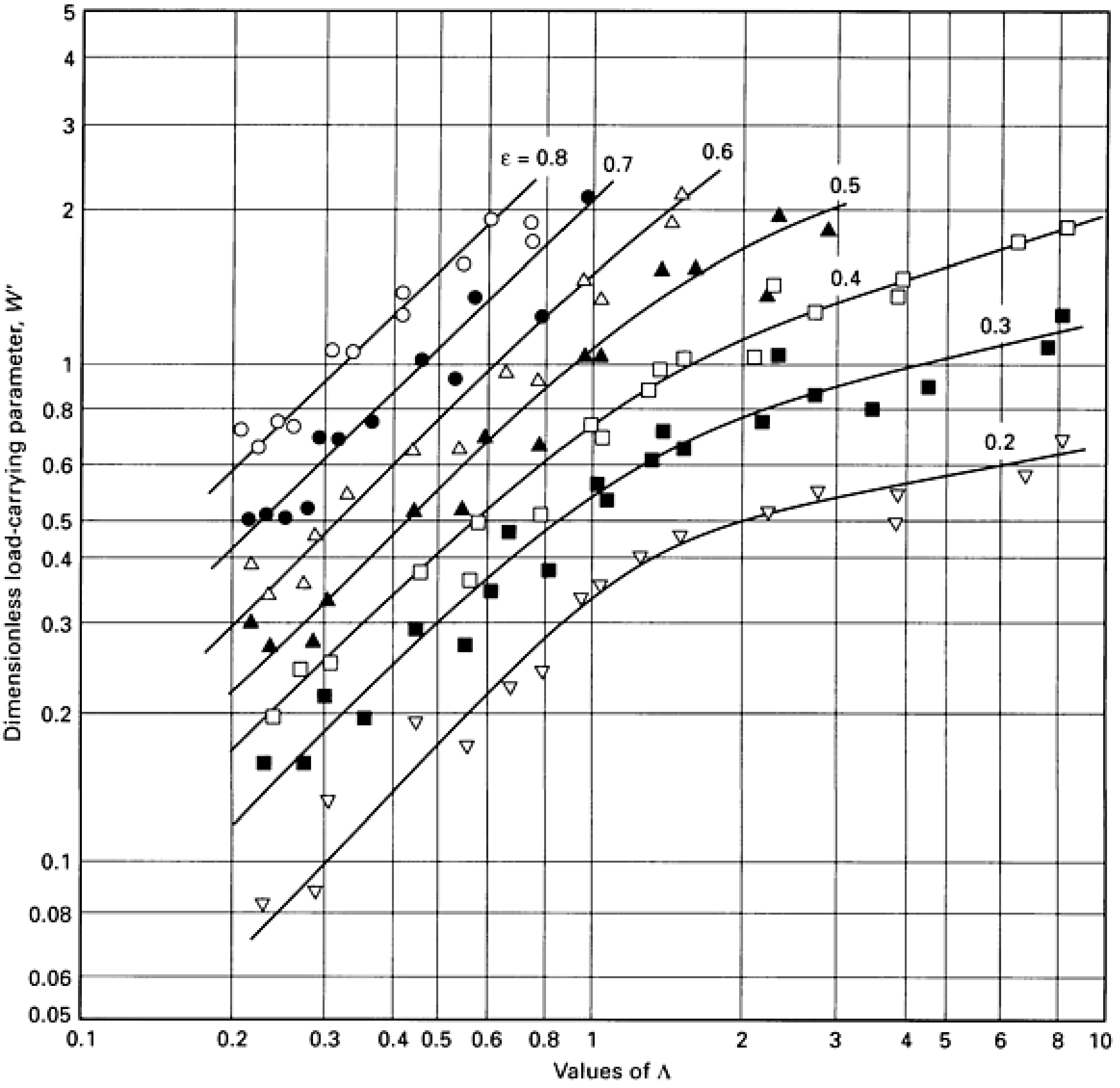
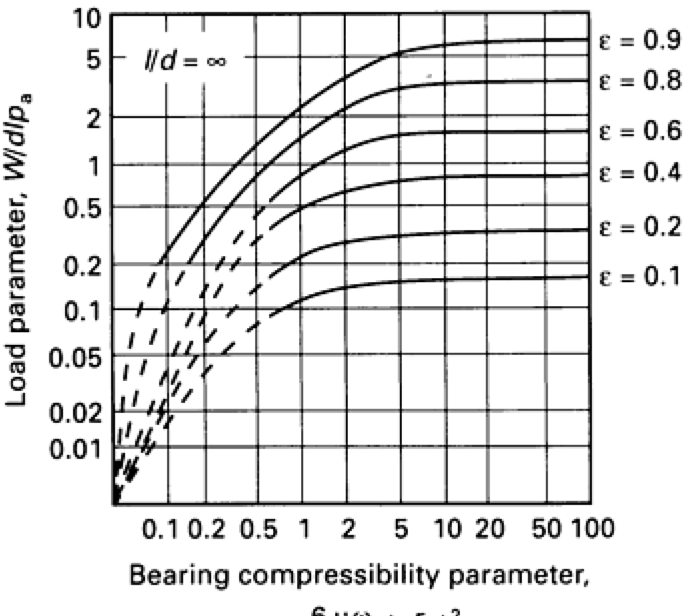


Fig. 14 Load parameter *W*' evaluated in terms of Λ and ϵ. Source: Ref 7

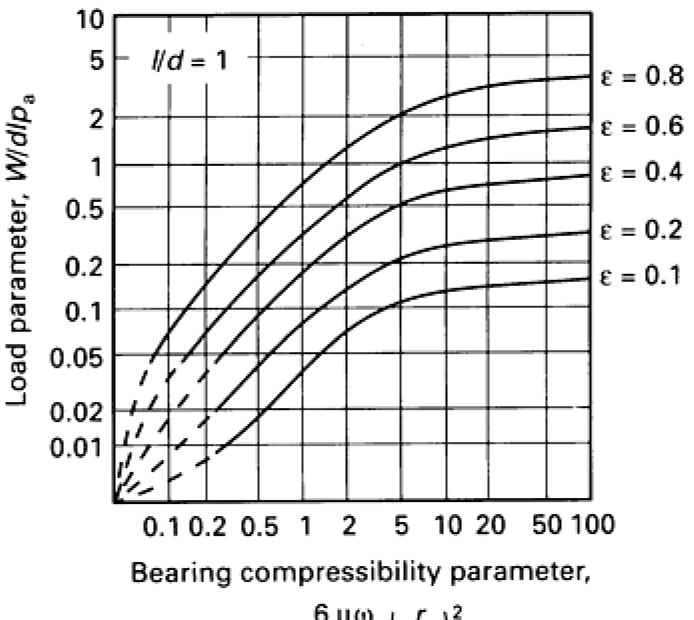
Another important variable in journal bearing analysis, especially stability analysis, is the attitude angle. In Fig. 1 this angle is shown as the angle, ϕ, between the line of load action, *W*, and the line of centers of the bearing, *O*-*O*'. The attitude angle will vary with load, speed, and other operating conditions, and also will depend on the geometrical configuration of the bearing itself. In general, if the attitude angle is small, the bearing is more stable against self-excited whirl. If the attitude is large, the bearing will be less stable against self-excited whirl.

For load-carrying capacity, the accumulated results of Gross (Ref 4), Elrod and Malanoski (Ref 15, 16), Elrod and Burgdorfer (Ref 17), and Raimondi (Ref 18) can be used as a general design approach for complete (360°) journal bearings. Figures 15, 16, and 17 show typical results from these analytical evaluations.



$$\Lambda = \frac{6\,\mu\omega}{p_a}\left(\frac{r}{c}\right)^2$$

Fig. 15 Theoretical load-carrying parameter versus compressibility number for full journal bearing with *l/d* = ∞



$$\Lambda = \frac{6\,\mu\omega}{p_a}\left(\frac{r}{c}\right)^2$$

Fig. 16 Theoretical load-carrying parameter versus compressibility number for full journal bearing with *l/d* = 1

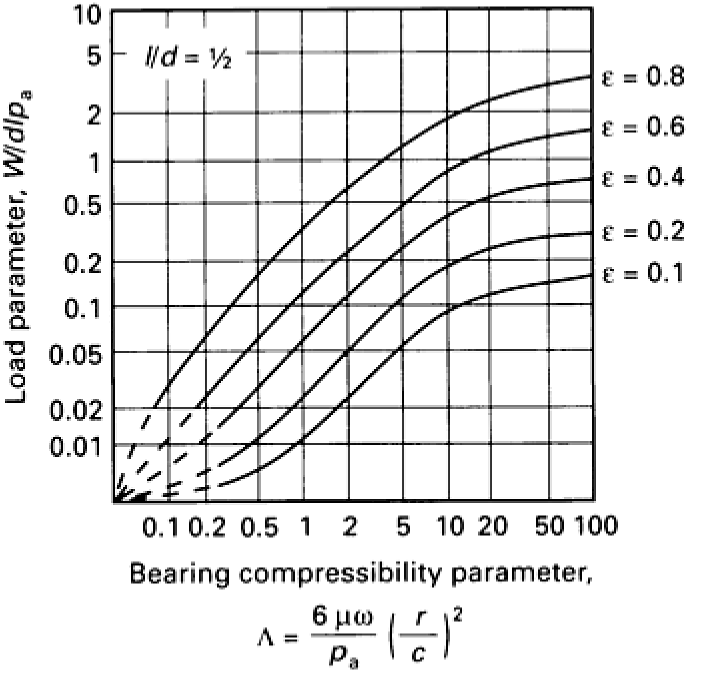


Fig. 17 Theoretical load-carrying parameter versus compressibility number for full journal bearing with *l*/*d* = $\frac{1}{2}$

Friction in Gas-Lubricated Journal Bearings

Of interest in some design applications is the power loss in the bearing due to friction resulting from viscous drag effects of the gas. Raimondi (Ref 18) has graphs of friction factors applicable to four *l*/*d* ratios, namely ∞, 2, 1, and $\frac{1}{2}$ for values of ϵ.

Specifically, the dimensionless friction factor for a plain journal bearing of ∞length is defined as:

$$C_f^{\infty} = \frac{M_J}{2\pi\mu U r^2 l/c} \qquad \textbf{(Eq 5)}$$

where,in English units, M_J is the frictional moment on the journal (in. · lbf). Figure 18 shows the friction factor plotted against the compressibility parameter.

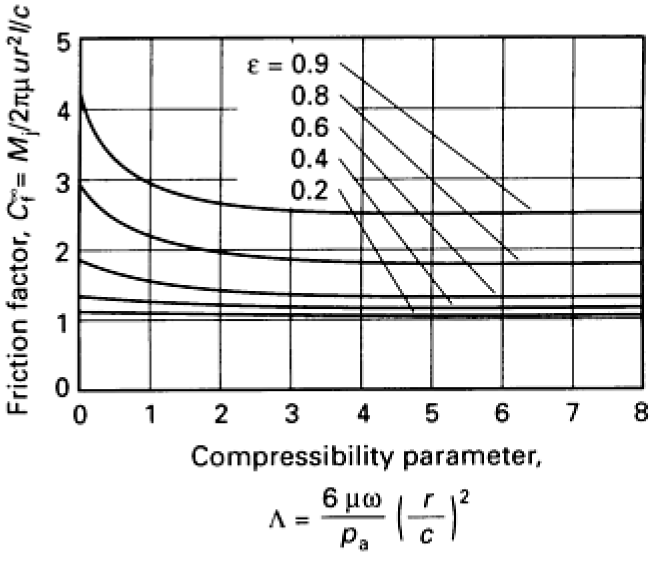


Fig. 18 Friction factor for full bearing of infinite length. Source: Ref 16

It can be seen from Fig. 18 that the friction loss is quite independent of Λ up to values of ϵ about 0.5. This means that for many bearings with low ϵ, values of friction can be reasonably estimated by using only simple Newtonian viscous shear. That would be shown by:

$$M_J = \mu A r \frac{v}{h} \quad \textbf{(Eq 6)}$$

where *A* is the swept area, which equals 2 πrl; *v* is the linear surface velocity; and *h* is the radial clearance, which equals *c*.

Synchronous Whirl

Synchronous whirl is a term used to describe a vibration condition in a fluid-film bearing where the mass of the rotor, combined with the spring stiffness of the bearing fluid film, can result in a natural frequency of vibration that is troublesome. If this natural frequency is the same as the rotational speed of the shaft, very large motions may be produced, resulting in possible bearing failure. If this natural frequency is below the operating speed, then it is necessary to pass through this resonant condition when approaching bearing speed.

Synchronous whirl can take the form of a translatory whirl, where the center line of the shaft describes the locus of a cylinder, or synchronous whirl can be conical. The conical mode form would resemble a bow tie, where there is no displacement at the center, only at the ends of the rotor. The displacement at the ends of the rotor would be out phase with each other.

If the bearing housing itself is elastically supported, it may also be capable of vibrating in either translatory or conical modes of vibration. These phenomena are analyzed in considerable detail in Ref 9.

A generally satisfactory approach is to estimate the quasi-static stiffness of the bearing film and then, in conjunction with the mass of the rotor, evaluate the natural frequency of vibration. It is a method that has proven its value over dozens of applications; Ref 9 provides examples of analyses that demonstrate the procedure.

Tilting-Pad Journal Bearings

One favored bearing design to achieve rotor stability is to use tilting-pad journal bearings. Considerable analytical and experimental work has been devoted to these bearings, both in the liquid- and gas-lubricated forms. These bearings also offer circumferential and axial misalignment capabilities.

Placing a resilient support (like a spring) under one pad of a three-pad bearing permits accommodation of centrifugal growth of the shaft at high speed or thermal or elastic dimensional changes in both shaft and bearing housing. When properly designed, this bearing is very forgiving. Many succesful applications have been made.

When using gas as the lubricant, some designs will include an auxiliary hydrostatic lift for starting (Fig. 20), which will reduce initial solid contact friction and protect the surfaces. The complete design is quite complex and an area of great specialization. Design guides for gas-lubricated journal and thrust bearings can be found in Ref 21, 22, 23, 24, and 25.

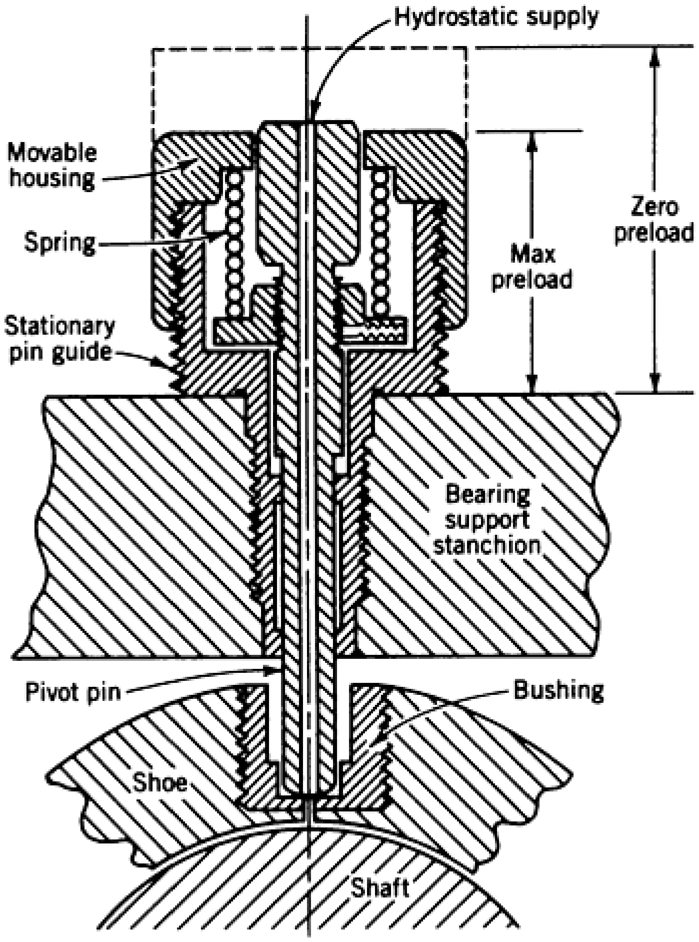


Fig. 20 Cross-sectional view of a spring-mounted pivot assembly. Source: Ref 21

A few fundamental concepts are critical in the design of a tilting-pad journal bearing. Of primary importance are yaw stability, pivot circle clearance, and pivot design. Other design features--such as film stiffness of combined bearing, minimization of bearing friction, the effect of clearance on minimum film thickness, shoe pitching frequency, and rotor critical speeds--are examined in the referenced design manual (Ref 21).

Yaw Stability. A pivoted shoe (or pad) has three degrees of freedom. It can pitch and roll like a ship, and it can yaw. To yaw is to act like a boat rotating about a vertical axis through its center of gravity. Any such rotation will lead to film rupture at the outer corners of th pad, with resulting solid contact. To provide for yaw stiffness in the bearing pad, the angle of wrap of the shoes should be no less than 90° of circumferential arc. With shorter arcs, the pad rapidly loses its ability to provide a resisting moment.

Pivot Circle Clearance. One of the crucial parameters governing the operation of the bearing is the setting of the pivot circle clearance (C' in Fig. 21). If the shaft always remained in the central concentric position, then C' could equal the actual clearance in the bearing, designated by C. That is to say, the radius of curvature of the shoe equals the radius of curvature of the shaft, R, plus the clearance C. The radius of curvature of the shoe is thus greater than the radius of curvature of the shaft. This produces the same phenomenon as having a crowned tilting pad. The load-carrying film cannot collapse even if the film becomes very thin.

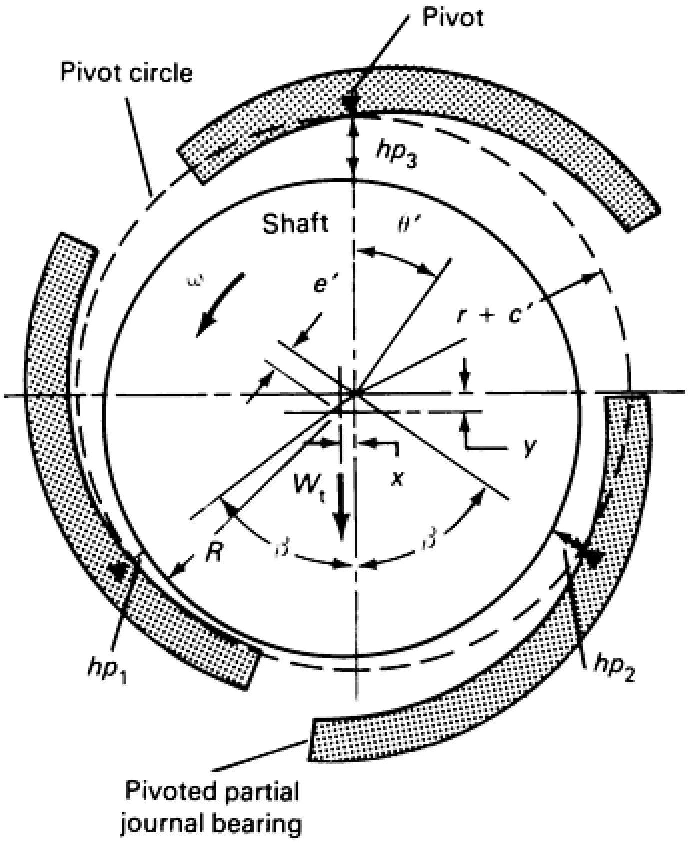


Fig. 21 Pivoted-pad journal bearing with three shoes. Source: Ref 21

However, with the journal bearing under load, and with low speeds, the eccentricity of the shaft can become great enough so that it moves away from the top (unloaded) tilting pad, so that the pad will lose its converging wedge and become unstable. To avoid this possibility, the shaft is preloaded by squeezing the pads radially inward, making C' less than C. The measure of this is the ratio C'/C, which should be less than 1. This is a design parameter and depends on the expected range of eccentricity ratios to be encountered.

Pivot Design. The third important concept is the design of the actual pivot. A ball and socket joint will not be satisfactory because of the frictional resistance to sliding. As the pad assumes its equilibrium position, its righting capacity goes to zero. Any residual friction in the pivot will prevent the pad from ever reaching its equilibrium design inclination. Figure 22 shows a recommended configuration utilizing rolling friction rather than sliding friction. Incidentally, the surfaces should be smooth and very hard to prevent fretting corrosion.

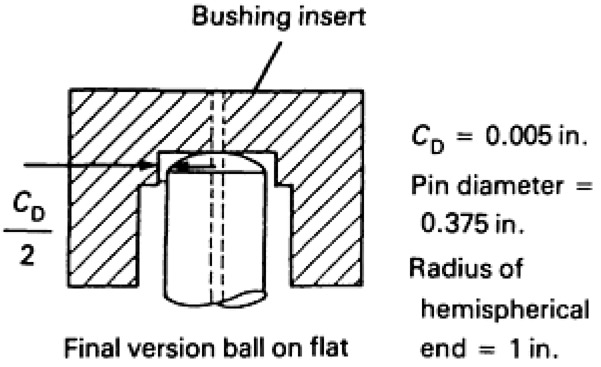


Fig. 22 Bushing insert for pivoted-pad hydrostatic journal bearing. Source: Ref 21

Three-Sector Journal Bearing

This journal bearing, with three equally spaced axial grooves, has shown some degree of stability against self-excited whirl. Castelli and Pirvics (Ref 26) have presented comprehensive numerically computed performance characteristics for three- and four-axial grooved gas-lubricated journal bearings. Angular extent of each groove is considered to be 5°. Thus for a three-sector bearing, the arc length of a sector would be 120 - 5 = 115°.

Figure 23 lists results in terms of the dimensionless load capacity parameter $W' = W/p_a\ rl$ plotted against the bearing compressibility number Λ. Note that the value of Λ used here is actually the same as in Eq 1. The applied load is directed toward the center of one sector. Evaluating W' and Λ will determine the eccentricity ratio and thus the minimum film thickness.

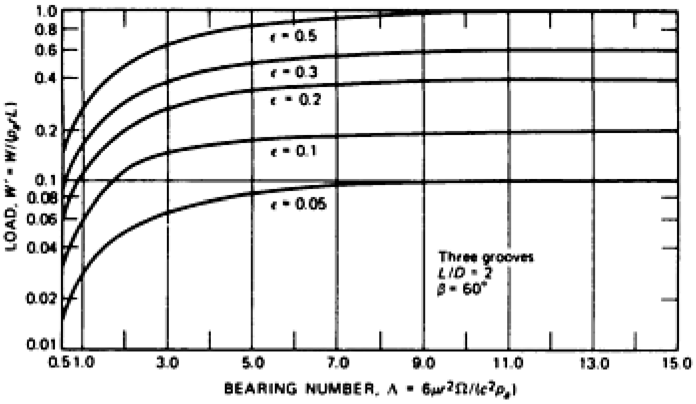


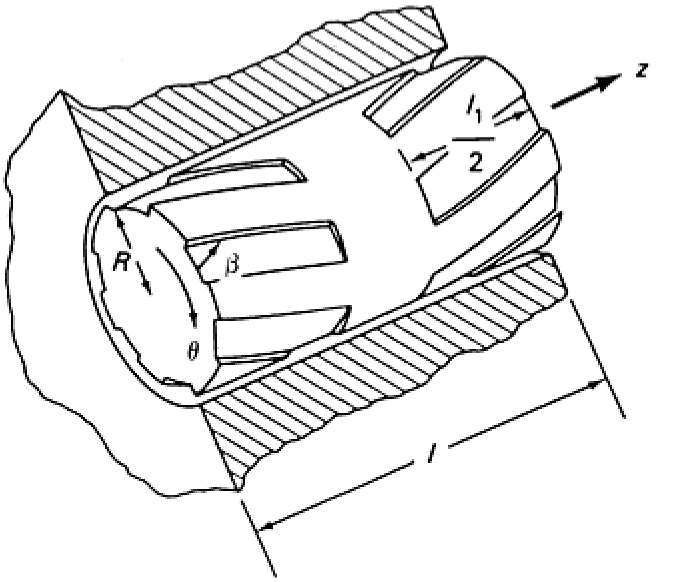
Fig. 23 Load function W' versus Λ for three-groove journal bearing. Source: Ref 26

Information on many other sizes, load directions, and attitude angles can be found in Ref 26 and 27.

Helical-Grooved Journal Bearings

The helical grooving in this type of journal bearing enhances stability by reducing the attitude angle below that obtained from a plain cylindrical journal bearing. These bearings are known for their stability and are often used as a possible substitute for tilting-pad journal bearings.

Castelli and Vohr (Ref 28) solved the appropriate equations numerically for load capacity and attitude angle for the case of l/d = 1.0, with various values of Λ. Figure 24 lists the geometric parameters for the spiral-grooved bearing as used in Ref 28. Figure 25 lists the results showing the load parameter $W' = Wp_a$ as a function of Λ with the eccentricity ratio ϵ as the third variable. Malanoski (Ref 29) shows good comparison between the theoretical predictions of Castelli and Vohr and his own measured results for helical-grooved journal bearings.



$$\alpha = \frac{a_g}{a_g + a_r} \qquad \bar{z} = \frac{l_1}{l} \qquad \Gamma = \frac{h_g}{h_r}$$

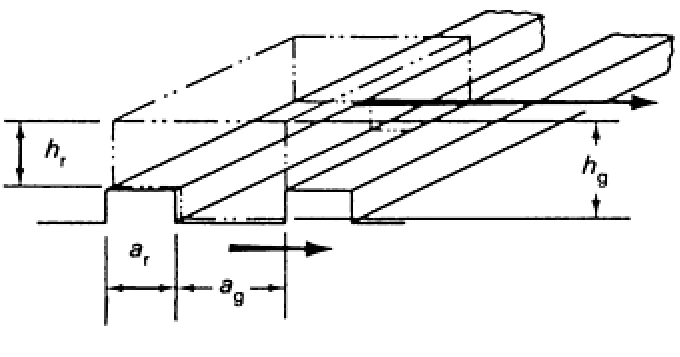


Fig. 24 Geometry of spiral-groove bearing, using notation of Castelli and Vohr (Ref 28)

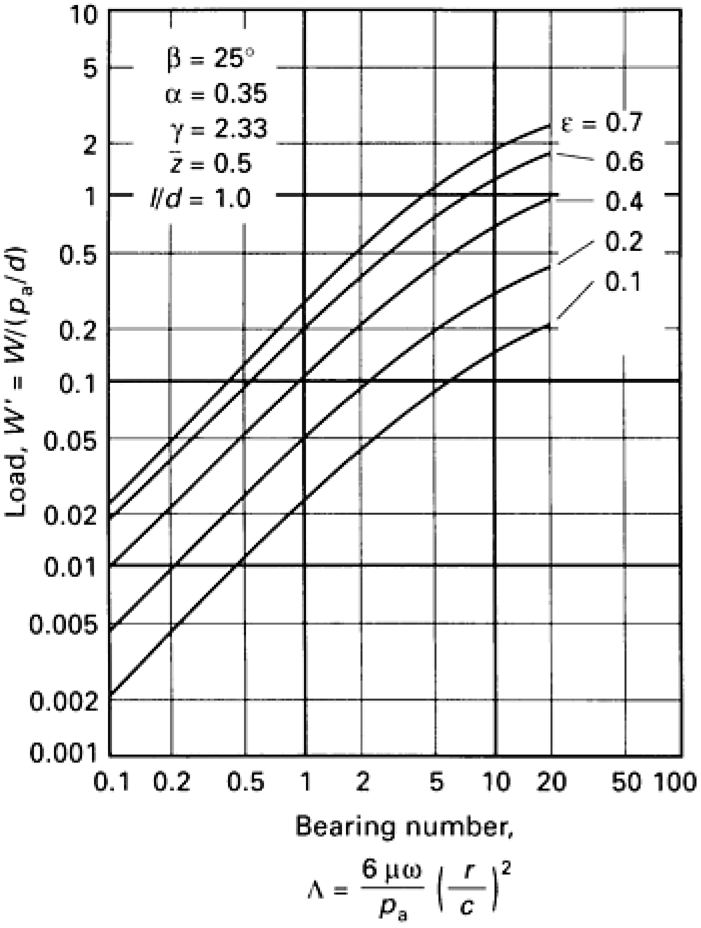


Fig. 25 Load capacity of spiral-groove journal bearing as a function of bearing number and eccentricity ratio. Source: Ref 28

Hydrostatic Gas-Lubricated Bearings

The many advantages of externally pressurized bearings are well known (Ref 9). With gas they have the added benefit of extreme cleanliness, and the use of gas enables them to operate over a wide range of temperatures.

However, analytical and design complications arise because of the compressibility of the gas. At low supply pressures (gage) equal to or less than ambient pressure (absolute), the system can be very simple; for example, supply pressure 70 kPa (10 psig) with an ambient pressure of 100 kPa (14.7 psia).

With high feed pressure, the gas flow in the entrance section is extremely complicated and may involve choked flow, shock waves, vortex formation, and boundary layer growth. Many comprehensive studies have been made of supersonic pressure depression in the feeding region of externally pressurized bearings (Ref 30, 31, 32, 33, 34).

These bearings do not involve a constant volume of flow as is the case with many liquid-lubricated hydrostatic bearings. Therefore, in order to achieve stiffness, they must have some kind of upstream restrictor in the feed line (Fig. 26). The flow restrictor can be an orifice or a capillary, and the bearing is then described as being restrictor compensated; the orifice restrictor area is equal to $\pi d_0^2/4$ (Fig. 26).

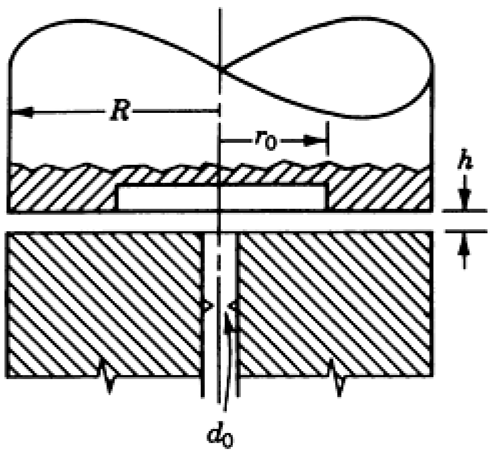


Fig. 26 Flow-restricted hydrostatic gas-lubricated bearing

Sometimes, the resistance to flow at the entrance to the film itself may dominate. In that case the bearing is identified as having inherent compensation. Many bearings are of this type. The inherent restrictor area would then be the circumferential annulus $\pi r_0 h$ at the entrance to the film (Fig. 26).

A typical pressure profile is shown in Fig. 27 for a simple circular thrust bearing with a central feed source. With a supply pressure of 480 kPa (70 psig) and a film thickness on the "sill" of h = 0.05 mm (0.002 in.), the effect of sonic velocity is seen. The minimum pressure measured on the sill (p_s) is 76 kPa (11.01 psia), indicating a partial vacuum. The maximum recovery pressure on the sill is 101 kPa (14.6 psig).

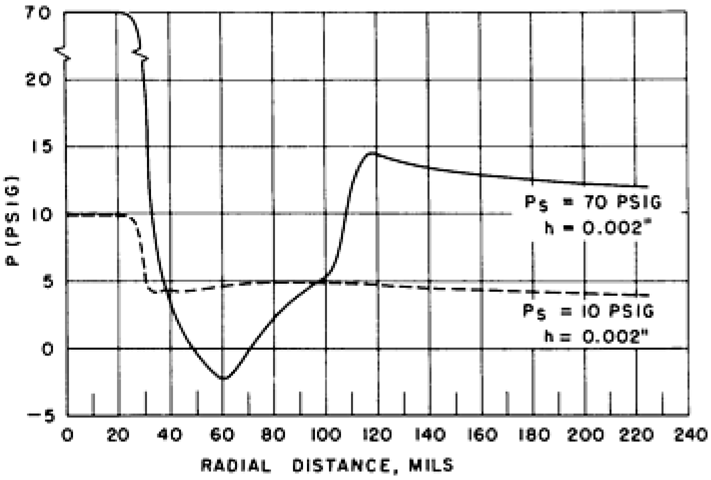


Fig. 27 Experimental pressure profiles for simple area 2 $\pi r_0 h$ externally pressurized thrust bearing

However, notice in Fig. 27 that the radial distances are measured in mils (0.001 in. = 0.025 mm), so that all of this "micro-aerodynamic" activity has taken place within a radius of 120 mils, or about 3.2 mm. The remainder of the bearing area can thus be treated as laminar isothermal gas flow and analyzed accordingly.

The dashed line in Fig. 27 shows the pressure profile when the bearing feed pressure was 70 kPa (10 psig). There is no sonic flow. Allowing for the pressure drop in the restrictor, either external or inherent, the bearing characteristics can be obtained in a relatively simple manner.

Figure 28 is a sample of the excellent program conducted by Laub (Ref 35) on thrust bearings and journal bearings intended for metrology applications. The pressure profiles show no indication of significant flow restriction and localized pressure loss because of bearing geometry and low pressure levels that were used. Flow evaluation, however, must include compressibility effects.

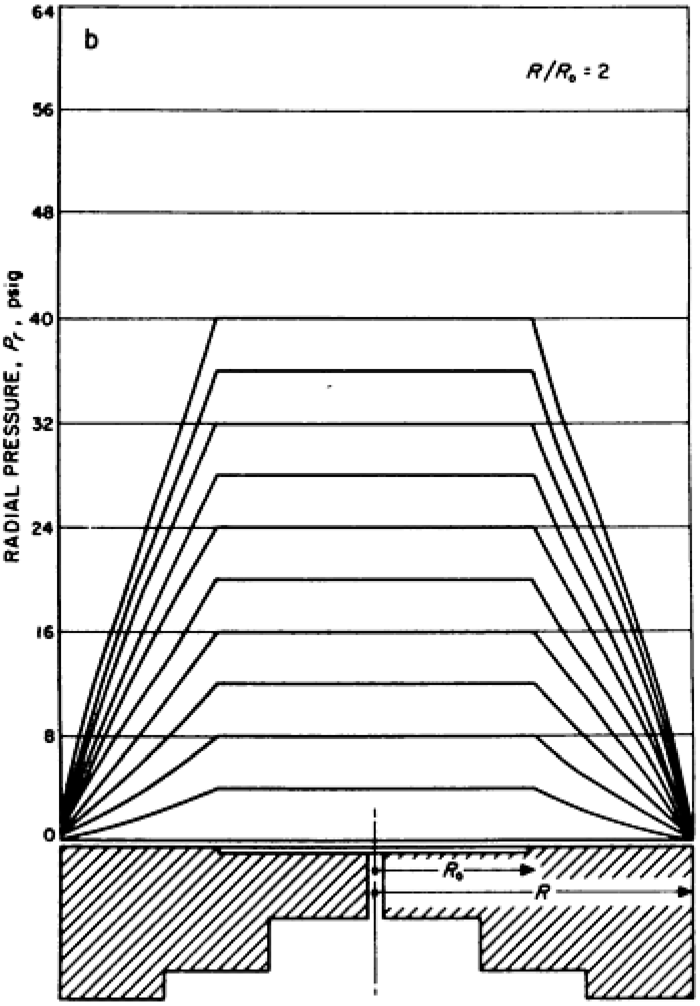


Fig. 28 Pressure profiles in gas-lubricated hydrostatic bearing: Source: Ref 35

As an example of the need to include the compressibility effect in flow prediction, consider the hydrostatic step bearing shown in Fig. 29. For liquids (incompressible), the supply pressure, P_0 and the ambient pressure, P_1, are gage pressures, R_0 is the radius of the recess, and R is the radius of the shaft. The derivation for the flow of lubricant in such a bearing is given in Ref 9 as:

$$Q = \frac{P_0 \pi h_0^3}{6\mu \ln (R/R_0)} \quad \textbf{(Eq 8)}$$

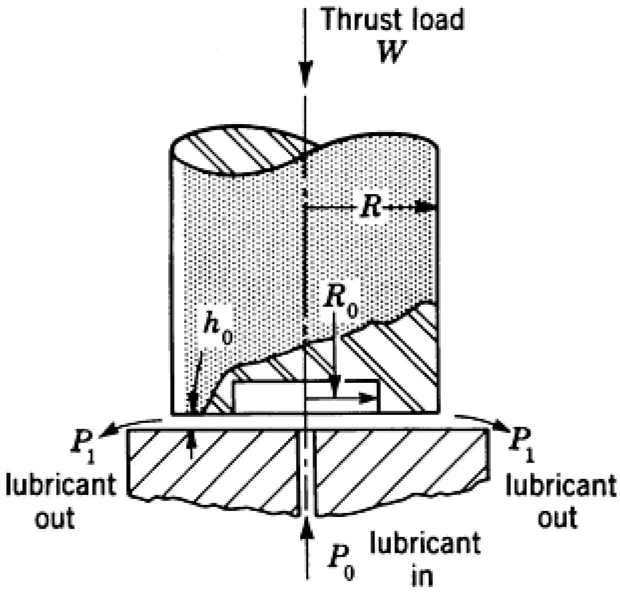


Fig. 29 Schematic diagram of a step bearing

However, when compressibility effects are included, the equation for flow volume becomes (Ref 9):

$$V_0 = \frac{\pi h_0^3}{6\mu \ln \frac{R}{R_0}} \left[\frac{P_0^2 - P_1^2}{2P_0} \right] \quad \textbf{(Eq 9)}$$

where P_0 and P_1 are in this case absolute pressures, and V_0 is the flow volume at the supply pressure P_0.

Pneumatic Hammer

The most troublesome characteristic of the externally pressurized air bearing is instability. During test programs, the phenomenon of self-excited vibration is often encountered, characterized violent fluctuations of pressure in the recess and amplitudes of vibration many times greater than the gap width at the equilibrium point. This phenomenon is often called pneumatic hammer.

Licht, Fuller, and Sternlicht (Ref 36) used a simplified lumped-parameter analysis to examine this problem. The gas film density in an oscillating thrust bearing is time dependent, and, in general, the mass inflow does not equal the mass outflow. As a consequence of film compressibility, energy from the film may be periodically added to the system in phase with the motion so that instability develops. The vibration is independent of system resonances.

The general stability analysis reveals the following:

- For a constant supply pressure, p_s, stability is enhanced by increasing the recess pressure p_0--that is, minimizing the pressure drop through the supply restrictor so that p_s - p_0 is a minimum. This of course reduces the stiffness of the bearing
- A recess depth comparable to the film thickness would be the ideal
- Maximizing the size of the inlet supply orifice will increase stability because capillary restriction is more likely to be unstable than orifice restriction
- Incompressible films are always stable

Multiple-Pressure Sources

To avoid pneumatic instability, it is clear that the high-pressure recess should be very shallow. The limit would be a recess of zero depth or no recess at all. However, that would reduce the load capacity of the bearing. A frequently used alternative is a ring or other appropriate pattern of multiple supply orifices that acts to develop an equivalent high-pressure area with the same depth as the film itself. Figure 30 shows a simple thrust bearing modified in this manner.

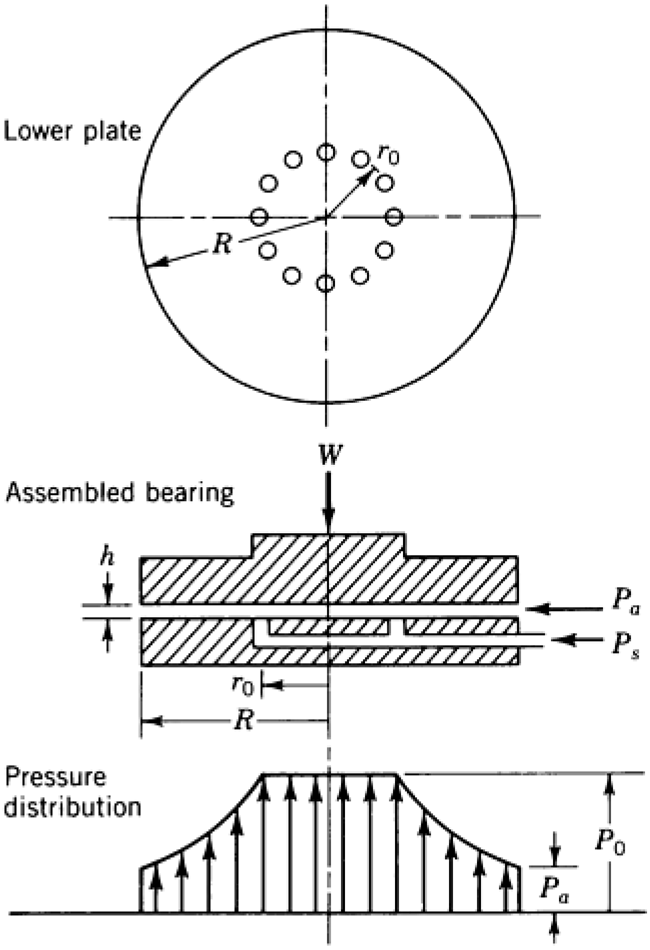


Fig. 30 Multiple-source feed for thrust bearing

In this bearing, the orifices are located on a circle of radius r_0. All orifices feed air to the interface at the same pressure P_0. Because there is no pressure gradient between the orifices, there is no flow between the orifices, and the entire circle of radius r_0 acts as a high-pressure recess. The same concept has also been successfully applied to journal bearings by Laub (Ref 35).

Porous Bearings

An alternate means for reducing the size and depth of a pressurized recess, other than using a finite number of multiple orifices as just described, is to feed the bearing through a section of porous material. Gas is admitted to the bearing interface through the pores of the material, resulting in a very large number of feeding restrictors in parallel. Again, the recess has been eliminated and stability enhanced. Many porous bearings have been made in both flat thrust and cylindrical journal bearing configurations. Sneck (Ref 37) provides an excellent survey of the many applications that have been made of this type of bearing, plus a very complete list of references.

Figure 31 is typical of a circular thrust bearing with pressurization through a porous annulus.Frequently, a porous carbon graphite is used so that antiscuff protection is provided by the material when in solid contact. A reasonable range of

permeability is available in these commercial products, and they have proven to be satisfactory. Clean air is essential to keep the pores from clogging with dust.

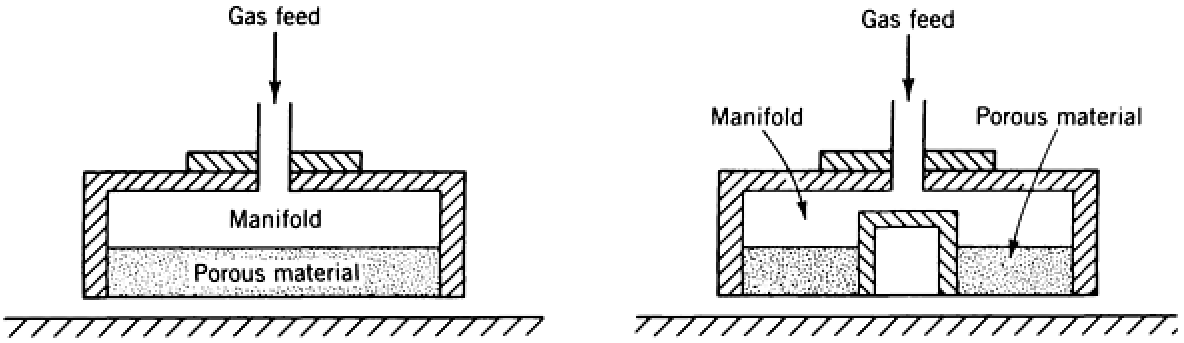


Fig. 31 Typical configurations for externally pressurized porous gas-lubricated thrust bearings. Source: Ref 9

Typically, pressurized porous bearings can be used anywhere classical orifice-compensated bearings are used. Design charts have been prepared by Gargiulo and Gilmour (Ref 38) to assist in a more exact analysis of these bearings.

However, it must be expected that the actual permeability of the material will be subject to some variation, even when cut from the same block. Vohr (Ref 4) discusses many additional design details involved in the use of porous materials in externally pressurized gas-lubricated bearings. These bearings can be extremely useful and are an attractive option for applications that call for a hydrostatic gas bearing.

Compliant-Surface Bearings

Compliant-surface bearings have been mentioned before in this article. They can use elastomers as the bearing material; in this form they have remarkable low-speed fluid-film capabilities. Foil bearings fall into this category as well. Compliant-surface bearings can be used as flexible membrane bearings.

The advantages of compliant-surface bearings include:

- Freedom from precision machining and maintenance of close tolerances
- Ability to accept misalignment
- Tolerance of dirt and particulates
- Accommodation of surface roughness with low surface speeds

The foil bearing (Fig. 32) is the most widely used form of compliant-surface bearing. It was first introduced, in simple form, by Blok and Van Rossum (Ref 39).

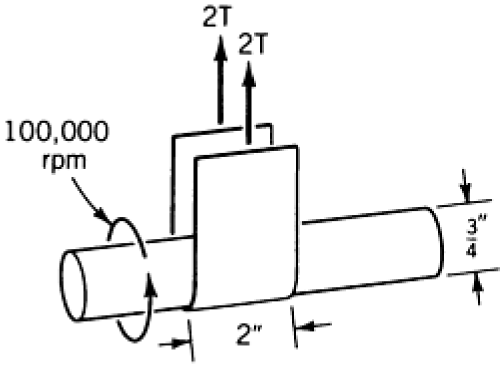


Fig. 32 Schematic diagram of a foil bearing. Source: Ref 27

The bearing can consist of a thin strip of flexible material (such as a plastic tape or thin metallic foil) partially wrapped around simple journal like the saddle belly band on a horse. As the journal spins, a reasonably large force can be supported by the self-acting hydrodynamic film in the contact area between the tape and the journal.

The simple foil bearing was further analyzed by Patel and Cameron (Ref 40) in 1957. It was utilized in the United States as an air bearing for applying load to a rotating shaft as early as 1956 by Fischer, Cherubim, and Fuller (Ref 27). By far, the greatest value of this simple concept is in tape transport for high-speed magnetic tape recorders. In this application, the journal is stationary and contains the recording head for the read-out components, while the tape glides past. Foil bearing analysis and design are reviewed extensively in Ref 4 and 41.

Developments of the original foil bearing concept of Blok and Van Rossum have now reached the stage of commercial application. Their advantages are many. If a metallic foil is used, the bearing can operate at high temperatures, especially when lubricated with air or some other gas. There is no problem with the possible loss of clearance due to differential thermal expansion between shaft and bearing, as is often the case with rigid surface units. The foil bearing establishes its own operating film thickness at all times.

It can also tolerate misalignment. In manufacturing, the foil bearing greatly reduces the need for holding expensive dimensional tolerances. An additional benefit is its stability in conjunction with high-speed rotor applications. The foil bearing is often used just for this reason, because it effectively reduces the possibility of self-excited fractional-frequency whirl (Ref 42).

Three distinct commercial varieties of foil bearings are available:

- Tension-dominated foil bearings (Fig. 33a)
- Bending-dominated segmented foil bearings (Fig. 33b)
- Bending-dominated continuous foil bearings (Fig. 33c)

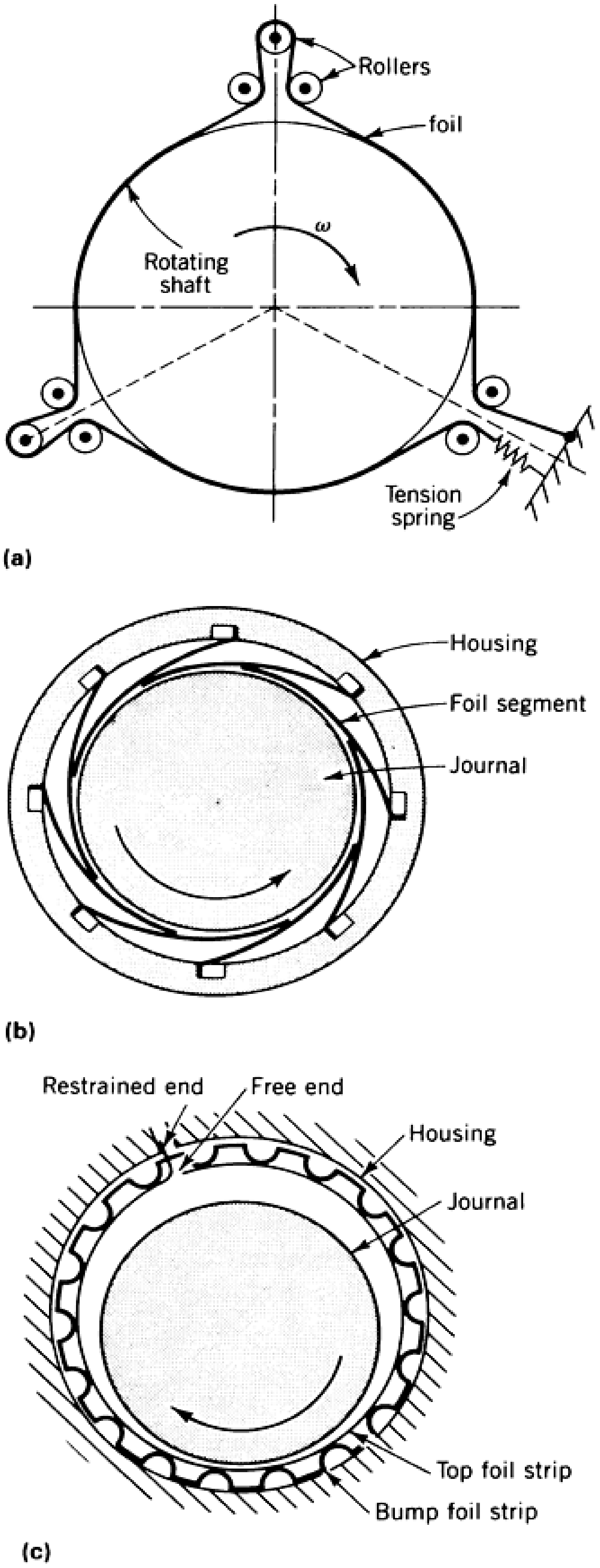


Fig. 33 Foil bearing designs. (a) Tension-dominated foil bearing. (b) Bending-dominated segmented foil bearing. Source: Ref 9. (c) Bending-dominated continuous foil bearing. Source: Ref 9

Tension-dominated foil bearings (Fig. 33a) have been used in data-processing equipment, where they provide support for magnetic tape traveling over guides, heads, and vacuum columns in tape transport. Recent developments have concentrated on adapting these bearings to the support of high-speed rotors for turbomachinery (Ref 43).

Bending-dominated segmented foil bearings of the type shown in Fig. 33(b) are widely used in both commercial and military aircraft applications. One example is the air-cycle refrigeration compressor used on many commercial transports for cabin cooling. Operating speeds can easily be 100,000 rpm, yet the compressor exhibits a high degree of reliability and freedom from rotor whirl instabilities. These bearings are being developed for cryogenic compressors at working temperatures near absolute zero.

A great deal of research and development work is being devoted to extending these bearings to high-temperature applications. With support from the Air Force Aero Propulsion Center and the Naval Air Propulsion Center, research programs are underway to develop a foil bearing capable of operating at high speed in a 650 °C (1200 °F) ambient. The potential application is for high-temperature gas turbines (Ref 44). Thrust bearing configurations are also used.

Bending Dominated Continuous Foil Bearings. This configuration has a single top foil strip that is restrained at one end and supported by a resilient corrugated foil strip called the bump foil strip (Fig. 33c). This piece elastically supports the top foil and controls the stiffness of the bearing. The deflection of the bump foil tolerates load fluctuations, and its elastic behavior provides resilience. Friction damping is also introduced with its beneficial stabilizing effects on the dynamics of the rotor.

Significant accomplishments have been recorded for bending-dominated continuous foil bearings, including:

- Three years of operation in automotive-type gas turbines at 60,000 rpm and 260 °C (500 °F)
- 120,000 rpm in product hardware
- Bearing shock resistance to 25 g at 100,000 rpm
- Tests at 650 °C (1200 °F)
- Unit bearing pressures for hydrodynamic air operation up to 345 kPa (50 psi), based on projected area under laboratory conditions

Modified Bending-Dominated Continuous Foil Bearings. This type of bearing is a modification of the design shown in Fig. 33(c). Although only limited applications have been made, it does appear to have excellent performance characteristics.

Design Analysis. There is no significant design analysis available, in the public domain, for the foil bearings described above. The design procedures are essentially proprietary. However, if a proposed application involves light starting loads ($\leq$7 to 14 kPa, or 1 to 2 psi) and when up to speed (>20,000 rpm) the unit loads do not exceed 70 to 100 kPa (10 to 15 psi), it appears that some type of foil bearing would be an excellent choice.

Pressurized-membrane bearings are bearings with a flexible membrane pressurized from the center through port holes (Fig. 34). At equilibrium conditions, the air escapes from under the inflated membrane through the minimum air gap. Levy and Coogan (Ref 45) first developed the analysis for such a device.

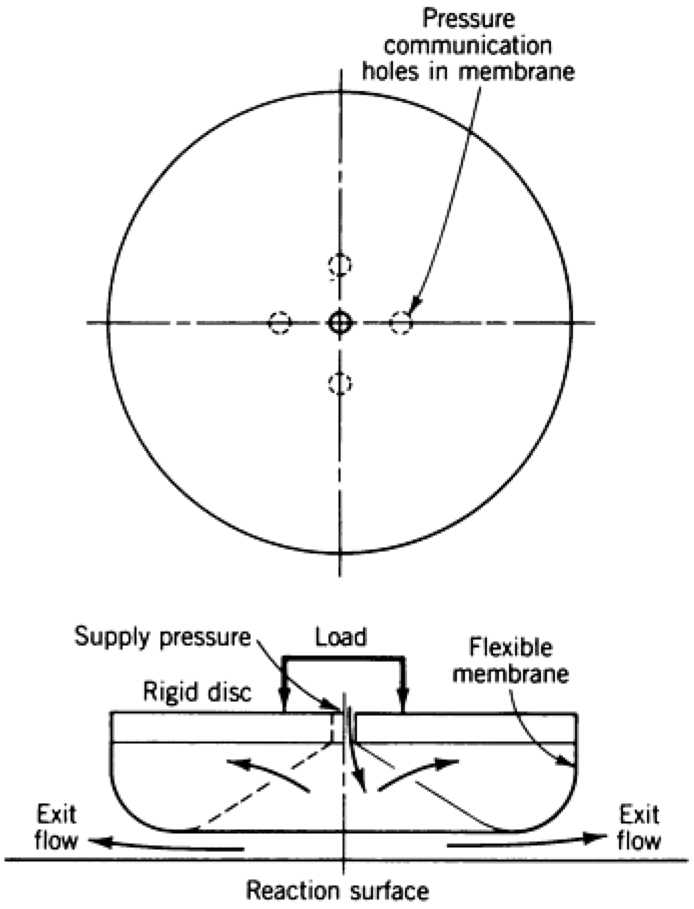


Fig. 34 Compliant-surface membrane bearing. Source: Ref 45

These bearings find wide use in areas where large loads must be lifted and transported over surfaces that may not be perfectly smooth, such as the floors in warehouses and factories. The flexibility of the membrane accommodates itself to the undulations that may exist, and motion proceeds smoothly with almost no friction. When the desired location or position is attained, the air supply is shut off and the object is set down. These devices are now available commercially.

A spectacular application of this type of hydrostatic membrane bearing is the Mile High Stadium in Denver, Colorado. The stadium was designed so that it would be suitable for both football and baseball. Accordingly, the entire side section of the grandstand, weighing 4080 Mg (4500 tons) and including 21,000 seats, is mounted at 46 points on water-lubricated hydrostatic rubber pads. Each pad is 1.2 m (4 ft) in diameter and made of fabric-reinforced synthetic rubber.

When pressurized, the entire section of the grandstand floats on a film of water and is moved back a distance of 44 m (145 ft) with very little effort. It is then set down in place, producing an arena suitable for playing baseball (Fig. 35). Although water was the pressurized lubricant of choice in this application, air could have been used.

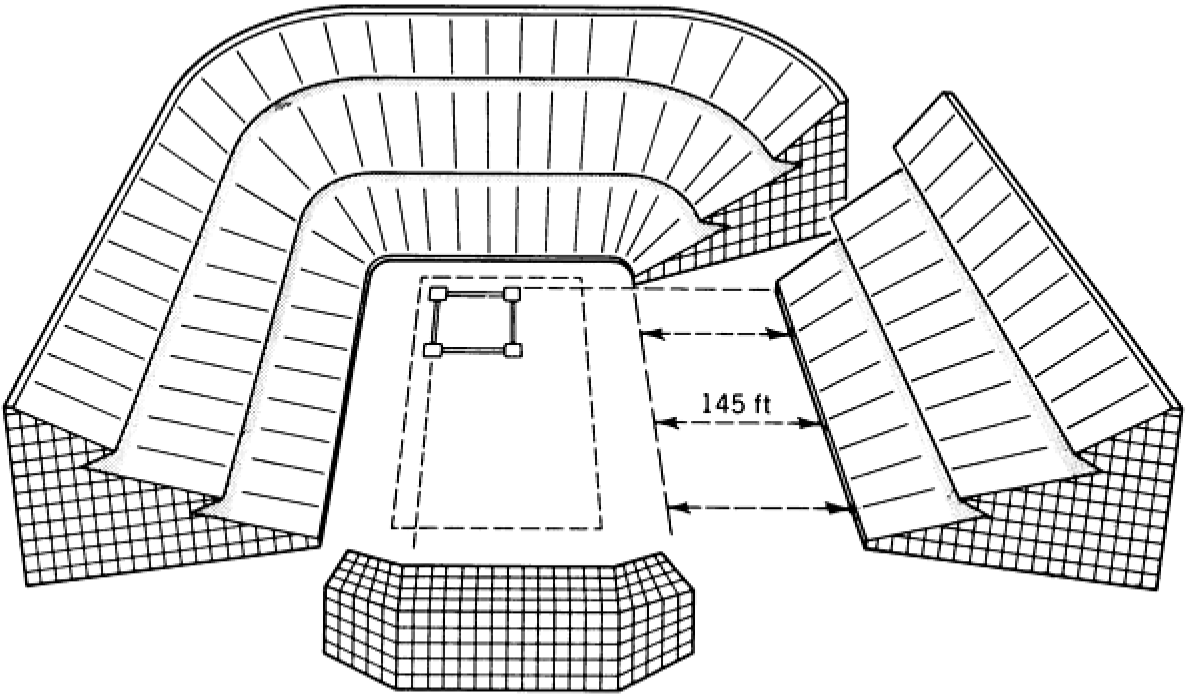


Fig. 35 Use of pressurized membrane bearings in Denver Mile High Stadium

It is surprising how little pressure is needed to support a large load. An available circular commercial unit 430 mm (17 in.) in diameter, with a pressure of about 21 kPa (3 psig) can carry a load of 270 kg (600 lb) with a flow of about 57 L/min (2 standard ft^3/min). The same unit at only 70 kPa (10 psig) can carry 900 kg (2000 lb) with, of course, a greater flow.

A novel application was made recently by a company in England that refurbishes full-size passenger railroad cars. A cradle platform was built to hold the four-wheel trucks at each end of the car. Then the platforms were floated on air-pressurized membrane bearings. Once this was done, the railroad car could be easily moved to any location on the factory floor without being encumbered by rails.

A few years ago, this type of hydrostatic compliant pad was offered as an option on a well-known brand of domestic refrigerator. By reversing the hose connection on a standard vacuum cleaner, the discharge pressure could be directed to the pads under the refrigerator and cause it to float in a film of air. It could then be moved very easily away from the wall, even when fully loaded, to allow for cleaning and maintenance.

Hydrostatic Journal Bearings

Externally pressurized journal bearings may be used when it is necessary to maintain a precise shaft position, with negligible friction, when rotational speed is insufficient to establish a hydrodynamic film. External pressurization is also used in combination with a hydrodynamic-type bearing to increase film stiffness, reduce the attitude angle, and raise the threshold of instability. The combination of these two actions produces what is known as a "hybrid bearing" (Fig. 36).

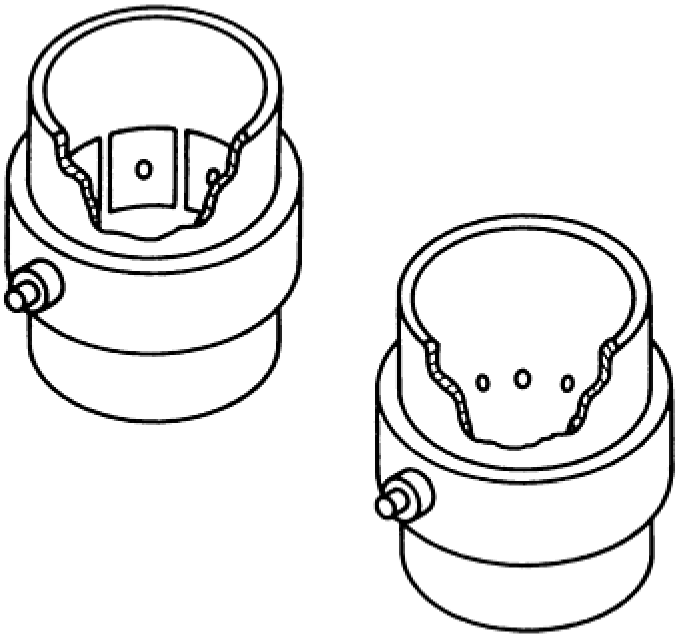


Fig. 36 Typical hybrid journal bearing designs. Source: Ref 9

The mutual contribution of each type of lubrication is of course of design decision. The rotor may be started or stopped without solid contact. Bearings for high-speed dental drills (500,000 rpm) are of the hybrid variety (Ref 46).

Externally pressurized journal bearings are affected by a great many combinations of parameters, so their design is relatively complex. Lund (Ref 47) has combined first-order self-acting and externally pressurized perturbation solutions for small eccentricity ratios. The results are found to be satisfactory and in good agreement with experimental results for small eccentricity ratios.

As with thrust bearings, the recesses must be kept as small as possible to avoid pneumatic instability. Multiple feed orifices or porous sections may be an answer to the problem. With multiple feeds, the assumption is usually made that the discrete points be considered as a continuous line source of pressure.

Comprehensive design charts for externally pressurized journal bearings are presented in Ref 48. Shapiro (Ref 49) has evaluated a three-sector hybrid journal bearing for both steady-state and dynamic behavior. He has published the results for this gas bearing, which uses orifice compensation, in tabular and graphical forms.

Many applications of externally pressurized gas bearings have been made to instruments, with remarkable increases in precision. Wunch, in Ref 45, describes the revolution that gas bearings have produced in the field of metrology. The surprising conclusion is that even though the components of the air bearing have been produced with normal manufacturing tolerances on dimensions and surface finishes, the rotational accuracy of the final instrument itself is of the order 0.075 to 0.125 μm. The explanation assumes the pressurized film has an averaging effect, and the small surface undulations of the bushing (surface roughness) are not transmitted to the shaft. This is similar to the smooth motion of a row boat on the surface of a pond; the boat is not influenced by the roughness of the bottom of the pond.

In machine tool applications, Lewis (Ref 50) has shown that grinder spindles, shapers, and gear cutters will actually show run-outs measured in microinches using externally pressurized journal bearings.

Materials for Gas-Lubricated Bearings

Engineers have been forced to search for materials that, when operated under extremely adverse conditions, will deliver acceptable life expectancy and reasonably low friction. These adverse conditions may be associated with high

temperature, with the use of lubricants with low viscosity and little or no natural oiliness (like gas), or with dry operation where no lubricants are present at all.

Experimentation has led in the direction of harder and harder materials, such as refractory materials and ceramics (Ref 51, 52, 53). These materials are usually brittle, so impact must be avoided. When they are used as rigid bearings with gas lubrication, the geometry must be precise and true because there is little possibility of conformability with such materials (as there could be with tin or lead-based babbitt, for example). Rigid designs, relatively light loads, and elimination of thermal or elastic deformations are necessary.

Investigations of gas-lubricated rigid thrust bearings and journal bearings operating under high temperature and high-speed conditions have shown favorable performance with the use of hard ceramic materials. Materials that might be considered for such applications include boron carbide, chromium carbide, silicon carbide, titanium carbide, alumina, silicon nitride, tungsten carbide, and chromium oxide.

Ceramics and cermets exhibit superior wear resistance. Cermets are ceramics that have been bonded with metals to improve their ability to handle impact and shock loading. The results, of course, vary with the ceramic and its bonding material. Cobalt is often used as a binder.

Hinkle and Fuller (Ref 54) conducted a study of the friction and wear of various materials for gas-lubricated rigid journal bearings under conditions of start-stop and high-speed whirl-induced rubbing. The tests were performed with a variety of imposed loads, temperatures, and ambient atmospheric conditions. Of the 36 combinations of materials investigated, by far the best performance was exhibited by aluminum oxide (alumina) against itself. The best specimens showed no visible evidence of wear even after extended testing.

Surprisingly, of the four different types of aluminum oxide used, one was far superior to the other three. The binder, the density, and the grain size are apparently strong variables. It might be advisable to get guarantees of friction and wear characteristics from a prospective vendor before selecting a particular grade of aluminum oxide.

Ceramics do provide low wear, but they also have relatively high friction. This may be a problem in some applications. In instrument gyroscopes, the electric motor drive is very small and light and has a low starting torque. Consequently, unless the friction of the bearing material is low enough, the motor will stall and not start. Much current research in gyro bearings is directed toward the reduction of friction in ceramic bearings through the introduction of some form of boundary lubrication.

Research carried out at the University of Rhode Island has determined that if final grinding of the ceramic is done in a bath of a prescribed boundary lubricant, a surface modification is achieved that provides lower friction. For silicon nitride, octadecanoic armide is recommended. For glass ceramics such as Pyroceram, the use of dioctedecyl disulfide is indicated.

In England, the British Royal Navy Scientific Service frequently uses silicon nitride against silicon nitride for gyro air bearings. They have been successful with small built-in graphite brushes that spread a thin film of solid graphite on the running surfaces. This technique not only reduces wear and lowers friction, but it also burnishes the surfaces to a high gloss while running. An alternative technique is burnishing the surfaces before installation with either molybdenum disulfide (MoS_2) or Teflon. Reducing friction with ceramics is an active research field.

Surface Coatings. High-density ultrahard coatings can be applied to the surfaces of ordinary commercial machine elements and provide them with a highly wear-resistant capability. Typical applications include parts for use in jet aircraft and in the chemical, textile, steel, and data processing industries. New coating technologies for tribological applications are described in Ref 55. Many different techniques can be used in making these surface depositions, depending on the end results desired.

Surface coating is a specialized field in which dramatic developments have been made within the last decade. For example, the metallic foil compliant-surface bearings, described earlier, have been coated with wear- and friction-resistant materials for use in various applications. Commercially bonded MoS_2, aluminum oxide, chromium oxide, and a three-component commercial blend of aluminum oxide, silicon oxide, and chromium oxide have all been used successfully for coating these bearings.

References

1. H.D. Smyth, *Atomic Energy for Military Purposes,* Princeton University Press, 1946
2. D.D. Fuller, Ed., *Proceedings, First International Symposium on Gas-Lubricated Bearings* (Washington, D.C.), ACR-49, 26-28 Oct 1959
3. *Proceedings, Second International Symposium on Gas Lubrication* (Las Vegas), 17-20 June 1968, American Society of Mechanical Engineers
4. W.A. Gross, Ed., *Fluid Film Lubrication ,* John Wiley & Sons, 1980
5. A.S. Raimondi, A Numerical Solution for the Gas-Lubricated, Full Journal Bearing of Finite Length, *Trans. ASME,* Apr 1961, p 131-155
6. W.A. Gross, A Gas Film Lubrication Study; Part I: Some Theoretical Analyses of Slider Bearings, *IBM J. Res. Dev.,* Vol 3, July 1959, p 237-255
7. G.W.K. Ford, D.M. Harris, and D. Pantall, Principles and Applications of Hydrodynamic-Type Gas Bearings, *Proc. Inst. Mech. Eng.,* Vol 171, 1957, p 93-113, discussion, p 113-128
8. O. Reynolds, On the Theory of Lubrication and Its Application to Mr. Beauchamp Towser's Experiments, Including an Experimental Determination of the Viscosity of Olive Oil, *Philos. Trans. R. Soc. (London) A,* Vol 177, Part I, 1886, p 157-234
9. D.D. Fuller, *Theory and Practice of Lubrication for Engineers,* 2nd ed., John Wiley & Sons, 1984
10. W.A. Gross, A Gas Film Lubrication Study; Part I: Some Theoretical Analyses of Slider Bearings, *IBM J. Res. Dev.,* Vol 3, July 1959, p 248
11. S. Abramovitz, Theory for a Slider Bearing with a Convex Pad Surface; Side Flow Neglected, *J. Franklin Inst.,* Vol 259 (No. 3), Mar 1955, p 221-233
12. R.K. Brunner, J.M. Harker, K.E. Haughton, and A.G. Osterlund, A Gas Film Lubrication Study; Part III: Experimental Investigation of Pivoted Slider Bearing, *IBM J. Res. Dev.,* No. 3, 1959, p 260-274
13. A. Kingsbury, Experiments with an Air Lubricated Journal, *J. Am. Soc. Naval Eng.,* Vol 9, 1897, p 267-292
14. H.F. Brubach, Some Laboratory Applications for the Low Friction Properties of the Dry Hypodermic Syringe, *Rev. Sci. Instrum.,* Vol 18, May 1947, p 363-366
15. H.G. Elrod and S.B. Malanoski, "Theory and Design Data for Continuous-Film, Self-Acting Journal Bearings of Finite Length," Report I-A 2049-13, Franklin Institute Laboratories for Research and Development, Nov 1960
16. H.G. Elrod and S.B. Malanoski, "Theory and Design Data for Continuous-Film, Self-Acting Journal Bearing of Finite Length," (Supplement to Report I-A 1049-13), Report I-A 2049-17, Franklin Institute Laboratories for Research and Development, June 1962
17. H.G. Elrod and A. Burgdorfer, Refinements of the Theory of the Infinitely Long, Self-Acting, Gas-Lubricated Bearings *Proceedings, First International Symposium on Gas-Lubricated Bearings,* ACR-49, U.S. Government Printing Office, Oct 1959, p 93-118
18. A.S. Raimondi, A Numerical Solution for the Gas-Lubricated, Full Journal Bearing of Finite Length, *Trans. ASME,* Apr 1961, p 131-155
19. A.C. Hagg, The Influence of Oil-Film Journal Bearings on the Stability of Rotating Machines, *J. Appl. Mech. (Trans. ASME),* Vol 68, 1946, p A211-A220. Discussion, Vol 69, Mar 1947, p A77-A78
20. V. Castelli and H.G. Elrod, Solution for the Stability Problem for 360 Degree, Self-Acting Gas-Lubricated Bearings, *J. Basic Eng. (Trans. ASME),* Vol 87, Mar 1965, p 199-212
21. E.J. Gunter, J.G. Hinkle, and D.D. Fuller, "Design Guide for Gas-Lubricated, Tilting-Pad Journal and Thrust Bearings with Special Reference to High-Speed Rotors," NYO-2512-1, U.S. Atomic Energy Commission, Eng. Development Branch, I-A 2393-3-1, Contract AT 30-1-2512, Nov 1964
22. E.J. Gunter, J.G. Hinkle, and D.D. Fuller, The Effects of Speed, Load, and Film Thickness on the Performance of Gas-Lubricated Tilting-Pad Journal Bearings, *Trans. ASLE,* Vol 7, 1964, p 353-365
23. J. Corniglion, G. Kilmister, H. Woodley and A.D. Richards, Theoretical Design and Practical Performance of Tilting-Pad Gas Bearings, Paper No. 19, *Gas Bearing Symposium Proceedings,* University of

Southampton, United Kingdom, Apr 1967

24. C. Mech, Some Practical Performance Aspects of the Design of Gas-Bearing Blowers and Some Performances of Industrial Machines, Paper No. 16, *Gas Bearing Symposium Proceedings,* University of Southampton, United Kingdom, Apr 1967
25. J.H. Dunn, "Inspection of Two Brayton Rotating Units after Extensive Endurance Testing," Report TM X-73569, Lewis Research Center, National Aeronautics and Space Administration, 1976
26. V. Castelli and J. Pirvics, Equilibrium Characteristics of Axial-Groove Gas-Lubricated Bearings, *J. Lubr. Technol. (Trans. ASME),* Vol 85, p 177-195
27. G.K. Fischer, J.L. Cherubim, and D.D. Fuller, "Some Instabilities and Operating Characteristics of High-Speed Gas-Lubricated Journal Bearings," Paper No. 58-A-231, American Society of Mechanical Engineers
28. V. Castelli and J.H. Vohr, Performance Characteristics of Herringbone Grooved Journal Bearings Operating at High Eccentricity Ratios and with Misalignment, Paper No. 14, *Proceedings of Third Gas Bearing Symposium,* University of Southampton, United Kingdom
29. S.B. Malanoski, Experiments on an Ultrastable Gas Journal Bearing, *J. Lubr. Technol. (Trans. ASME),* Vol 89, Oct 1967, p 433-438
30. S.R. Carfagno and J.T. McCabe, "Summary of Investigations of Entrance Effects in Circular Thrust Bearings," Report I-2049-24, Franklin Institute Research Laboratories, Defense Documentation Center Report, AD619966, 1965
31. J.T. McCabe, H.G. Elrod, S. Carfagno, and R. Colsher, Summary of Investigations of Entrance Effects of Circular Thrust Bearings, Paper 17, *Proceedings of Fourth Gas Bearing Symposium,* Vol 1, University of Southampton, United Kingdom
32. P.S. Moller, Radial Flow without Swirl between Parallel Disks Having Both Supersonic and Subsonic Regions, *J. Basic Eng. (Trans. ASME),* Vol 88, 1966, p 147-154
33. J.H. Vohr, A Study of Inherent Restriction Characteristics for Hydrostatic Gas Bearings, Paper 30, *Proceedings of Fourth Gas Bearing Symposium,* Vol 2, University of Southampton, United Kingdom
34. H. Mori, A Theoretical Investigation of Pressure Depression in Externally Pressurized Gas-Lubricated Circular Thrust Bearings, *J. Basic Eng. (Trans. ASME),* Vol 83, 1961, p 201-208
35. J.H. Laub, Evaluation of Externally-Pressurized Gas Pivot Bearings for Instruments, *Proceedings, First International Symposium on Gas-Lubricated Bearings,* ACR-49, U.S. Government Printing Office, Oct 1959, p 435-481
36. L. Licht, D.D. Fuller, and B.Sternlicht, Self-Excited Vibrations of an Air-Lubricated Thrust Bearing, *Trans. ASME,* Vol 80, 1958, p 411-414
37. H.J. Sneck, A Survey of Gas-Lubricated Porous Bearings, *J. Lubr. Technol, (Trans. ASME),* Oct 1968, p 804-809
38. E.P. Gargiulo and P.W. Gilmour, A Numerical Solution for the Design of Externally Pressurized Porous Gas Bearings: Thrust Bearings, *J. Lubr. Technol. (Trans. ASME),* 1968, p 810-817
39. H. Blok and J.J. van Rossum, Foil Bearing--New Departure in Hydrodynamic Lubrication, *Lubr. Eng.,* Vol 9, 1953, p 316-320
40. B.J. Patel and A. Cameron, The Foil Bearing, Paper 73, *Proceedings, Conference on Lubrication and Wear,* Institute of Mechanical Engineers (London), Oct 1957, p 219-223
41. W.A. Gross, Analysis and Design of Foil Bearings, Paper No. 23, *Proceedings, Gas Bearing Symposium,* University of Southampton, United Kingdom, Apr 1967
42. F.J. Suriano, R.D. Dayton, and F.G. Woessner, "Test Experience with Turbine-End Foil-Bearing-Equipped Gas Turbine Engines," Paper 83-GT-73, American Society of Mechanical Engineers, 1983
43. L. Licht, The Dynamic Characteristics of a Turborotor Simulator on Gas-Lubricated Foil Bearings, *J. Lubr. Technol. (Trans. ASME),* Vol 94, 1972, p 211-222
44. "Gas Lubricated Foil Bearing Development for Advanced Turbomachines," Report AF APL-TR-76-114, Vol I and II, Air Force Aero Propulsion Laboratory, 1977
45. S.B. Levy and C.H. Coogan, Flexible Membrane Hydrostatic Air Bearing *J. Lubr. Technol. (Trans. ASME),*

Vol 90, 1968, p 184-190

46. N.S. Grassam and J.W. Powell, *Gas Lubricated Bearings,* Butterworths, London, 1964
47. J.W. Lund, The Hydrostatic Gas Journal Bearing with Journal Rotation and Vibration, *J. Basic Eng. (Trans. ASME),* Vol 86, June 1964, p 328-336
48. D.F. Wilcock, Ed., *MTI Gas Bearing Design Manual,* Mechanical Technology Inc., 1972
49. W. Shapiro, Steady-State and Dynamic Analyses of Gas-Lubricated Hybrid Journal Bearings, *J. Lubr. Technol. (Trans. ASME),* Jan 1969, p 171-180
50. T.G. Lewis, "Hydrostatic Gas Bearings and Metrology in Ultra-Precision Machining," 1965 ASME Spring Lubrication Symposium, American Society of Mechanical Engineers, June 1965
51. E.L. Hemingway, Surface Finish--Key to Bearing Life, *Mach. Des.,* Jan 1945, p 168, 170
52. J.T. Burwell, Jr., Teaching Lubrication and Bearing Design, Part V, *Mach. Des. Manuf. Bull.,* Vol XV (No. 6), March 1949
53. R. Poppinga, *Wear and Lubrication of Piston Rings and Cylinders,* Society of Tribologists and Lubrication Engineers (ASLE), 1948
54. J.G. Hinkle and D.D. Fuller, Evaluation of Friction and Wear Characteristics of Materials for Gas-Lubricated Bearings under Conditions of Start-Stop and Whirl-Induced Rubbing Paper No. 24, *Proceedings,* University of Southampton, United Kingdom, Apr 1967, p 24-1 to 24-34
55. W. Winer and M. Peterson, Ed., *Wear Control Handbook,* ASME Research Committee on Lubrication, American Society of Mechanical Engineers, 1980

Friction, Lubrication, and Wear of Gears

Robert Errichello, GEARTECH

Introduction

BECAUSE GEARS are such common machine components, they may be taken for granted. It is not generally appreciated that they are complex systems requiring knowledge from all the engineering disciplines for their successful design. Gear design is a process of synthesis in which gear geometry, materials, heat treatment, manufacturing methods, and lubrication are selected to meet the requirements of a given application. The designer must design the gearset with adequate strength, wear resistance, and scuffing resistance. To do this, he or she must consider gear tribology. The choice of lubricant and its application method is as important as the choice of steel alloy and heat treatment. The interrelationship of the following factors must be considered:

- Gear tooth geometry
- Gear tooth motion (kinematics)
- Gear tooth forces (static and dynamic)
- Gear tooth material and surface characteristics (physical and chemical)
- Lubricant characteristics (physical and chemical)
- Environmental characteristics (physical and chemical)

Note: Originally published in *Lubrication Engineering,* Jan-April, 1990. Reprinted by permission of the Society of Tribologists and Lubrication Engineers. All rights reserved.

Advantages and Disadvantages

The advantages of gas-lubricated bearings over liquid-lubricated fluid-film bearings are now well understood. These include:

- Cleanliness. Elimination of contamination caused by more traditional lubricants
- Reduction (often elimination) of the need for bearing seals
- Stability of lubricant. No vaporization, cavitation, solidification, decomposition, or phase change over extreme ranges of temperature, from cryogenic (-270 °C, or -450 °F) up to approximately 1650 °C (3000 °F). Operation at these extremes of temperature is a current research goal
- Low friction and heating with little or no cooling generally required. Permits practical attainment of high speeds (700,000 rpm)

Disadvantages of gas-lubricated bearings are recognized as resulting from the relatively low viscosity and damping of gas films. Thus, gas-lubricated bearings have a reduced load-carrying capacity compared to liquid-lubricated bearings, especially with self-acting or hydrodynamic bearings. For acceptable application, therefore, the bearings are necessarily larger and operate with thinner hydrodynamic films than their liquid-lubricated counterparts.

Thinner films demand closer control of manufacturing tolerances, surface finishes, and possible thermal and elastic distortions and alignments, to prevent rubbing contact. With compliant surface bearings, as with foil bearings and membrane bearings, rigid specifications regarding design and manufacture can be dramatically relaxed. The membrane bearing, for example, operates very satisfactorily over a typical factory floor.

The low damping of the gas film makes it necessary to carefully analyze the dynamic characteristics of the mechanical system employing the gas bearing, since if a critical speed or instability is encountered, there may not be enough damping

to suppress it or control it. With liquid-lubricated bearings these instabilities might not have been suppressed or passed over unnoticed because of the greater damping action that inherently exists with liquids. Much recent research has been devoted to the dynamics of gas bearings and their associated mechanical systems.

Gas-lubricated bearings have been characterized as being less forgiving than oil-lubricated bearings. This is certainly true for self-acting or hydrodynamic bearings. They are less forgiving of errors in estimating loads, of deviations from specifications during manufacture and installation, and of distortions and dirt that may afflict the rotor.

Compressibility Numbers

Gas is, of course, compressible and this effect must be included in the derivations for various forms of bearings--whether hydrodynamic (self-acting) or hydrostatic (externally pressurized). The extent of compressibility (represented by the compressibility number, Λ) is determined by a dimensionless group of parameters that actually evolves from the mathematical analysis. It takes on several forms depending on the geometry of the bearing.

For journal bearings (Fig. 1), the value of Λ is:

$$\Lambda = \frac{6\mu\omega}{p_a}\left(\frac{r}{c}\right)^2 \qquad \textbf{(Eq 1)}$$

where c is the machined-in radial clearance, p_a is the absolute ambient pressure, ω is the angular velocity of the journal (rad/s), μ is the absolute viscosity, and r is the radius of the shaft.

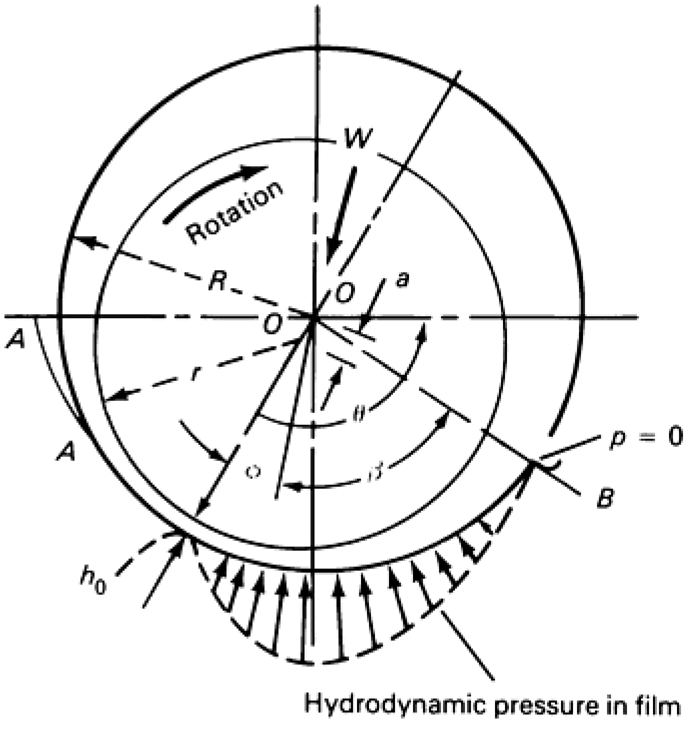


Fig. 1 Journal in full 360° bearing

For a tilting-pad thrust bearing (Fig. 2), the value of Λ is:

$$\Lambda = \frac{6\mu ul}{h_2^2 p_a} \qquad \textbf{(Eq 2)}$$

where l is the length of the pad in the direction of sliding (sometimes designated as L), and h_2 is the minimum film thickness (sometimes referred to as h_0), and u is the velocity of the runner past the shoe.

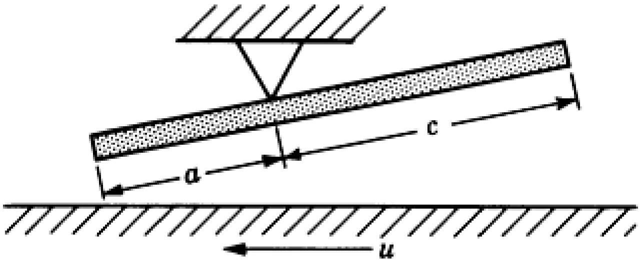


Fig. 2 Representation of pivoted shoe

For a Rayleigh step bearing (Fig. 3), the value of Λ is:

$$\Lambda = \frac{6\mu U l}{h_2^2 p_a} \qquad \textbf{(Eq 3)}$$

where l is the length of the pad in the direction of sliding, U is the velocity of the runner past the pad, and h_2 is the minimum film thickness.

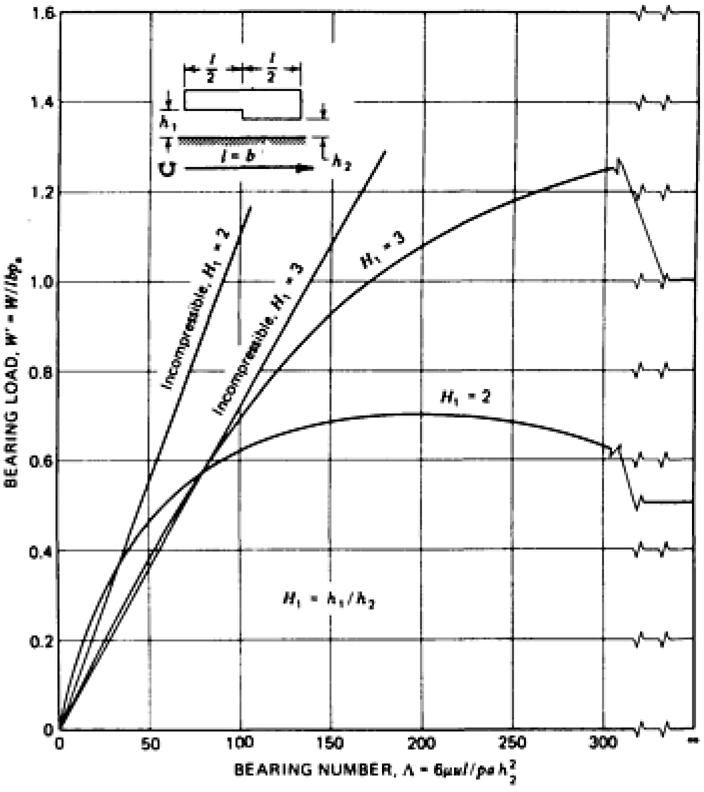


Fig. 3 Effect of bearing number on isothermal load for square-step slider bearings with film thickness ratios h_1/h_2 = 2 and 3. Source: Ref 4

When Λ approaches 0, operation of the gas-lubricated bearing approaches that of the liquid-lubricated (incompressible) case. As a Λ gets larger, as with lower ambient pressure or higher speed, the compressibility effects become very significant and must be included. For example, with journal bearings it can be shown that the equations for determining the load-carrying capacity for liquids and gases are essentially the same up to a value of Λ= 1 (Fig. 4). For tilting-pad bearings, the identity remain up to a value of about 15 (Fig. 5).

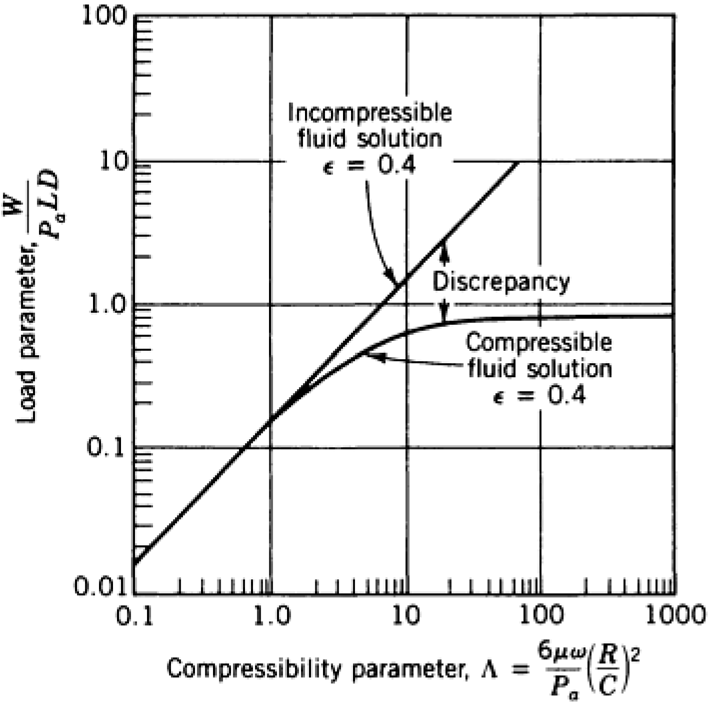


Fig. 4 Discrepancy in load capacity between results based on incompressible and compressible lubricants. Source: Ref 5

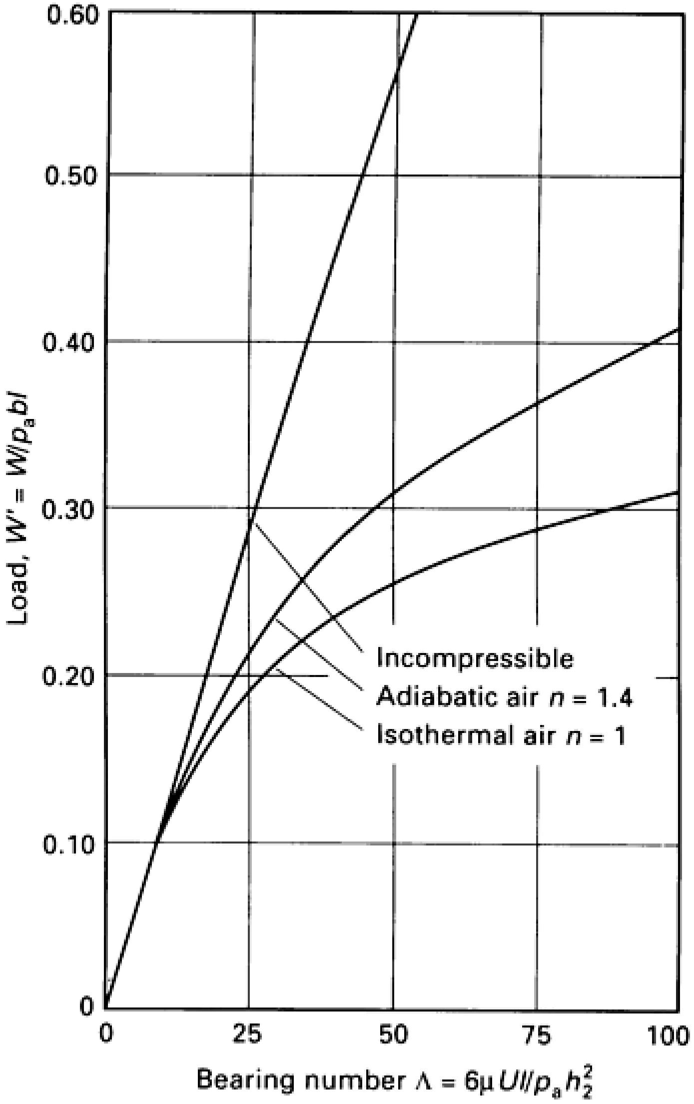


Fig. 5 Effect of bearing number on isothermal and adiabatic load for plane slider bearing operating in air with film thickness ratio h_1/h_2 = 2. Length-to-breadth ratio l/b = 1. Source: Ref 6

An interesting experimental demonstration is illustrated in Fig. 6, which shows the relationship between load-carrying capacity, *W*, and Λ for a hydrodynamic gas-lubricated journal bearing. Speed, viscosity, and eccentricity ratio were held constant while the ambient pressure was reduced, thus producing higher values of Λ. Notice that when the ambient pressure was reduced to 9% of atmospheric, the load-carrying capacity was lowered by 40%.

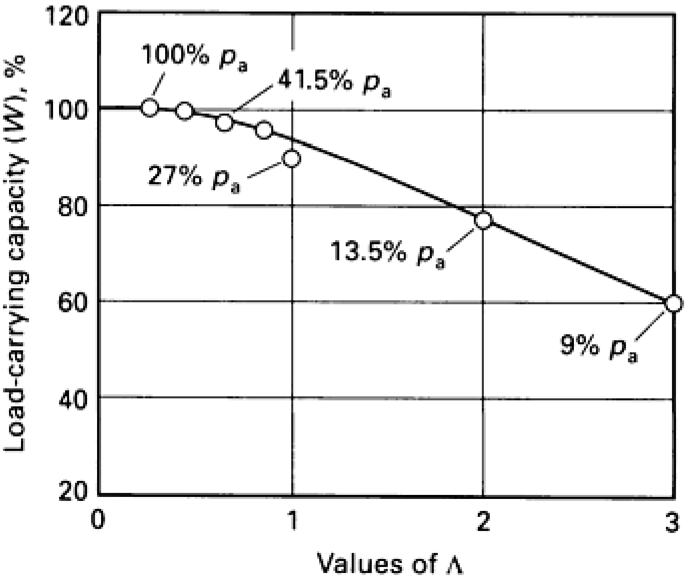


Fig. 6 Example of relationship between load-carrying capacity, *W,* and Λ for self-acting gas-lubricated journal bearing. Speed, viscosity, and eccentricity ratio held constant. Ambient pressure varied. Source: Ref 7

Definitions of Eccentricity Ratio and Clearance Modulus

In Fig. 1, *O* represents the center of the bearing and *O*' the center of the shaft. The distance *O*-*O*' is also designated as *a;* for light loads and high operating speeds, the center of the shaft and bearing coincide, and *a* approaches zero. The ratio of the distance *a* for any given operating condition to the machined-in radial clearance *c* of the bearing is called the eccentricity ratio, ϵ.

For heavy loads or extremes of operation, such as low speed or very low viscosity, the journal becomes more eccentric in the bearing and the distance *a* increases. The limit is where the journal just begins to make solid contact with the bearing, or the distance *a* equals the radial clearance. The eccentricity ratio is then 1. The radial clearance in the journal bearing is often designated as *mr,* where *m* is the clearance modulus of the bearing and *r* is the radius of the shaft or journal. The value of *m* will usually range between 0.0005 and 0.003 mm/mm, with typical industrial-type bearings running between 0.001 and 0.0025 mm/mm.

In the analysis of fluid-film bearings using incompressible lubricants, the continuity relationship was satisfied by saying that the volume of lubricant leaving the bearing was the same as the volume entering. With gas, of course, the volume changes, so that continuity must be based on equality of mass. Thus the mass of lubricant leaving equals the mass entering.

The basic Reynolds equation then becomes (Ref 8):

$$\frac{\partial}{\partial x}\left[\frac{\rho h^3}{\mu}\frac{\partial p}{\partial x}\right]+\frac{\partial}{\partial z}\left[\frac{\rho h^3}{\mu}\frac{\partial p}{\partial z}\right]=6U\frac{\partial \rho h}{\partial x} \quad \textbf{(Eq 4)}$$

where ρ is the mass density, *x* is the coordinate along the film in the direction of motion, *z* is the coordinate along the bearing dimension perpendicular to the direction of motion (sometimes called the length of the bearing, *l,* or the width of the bearing, *b*), *p* is the differential pressure, and *h* is the local differential film thickness.

Except for a few restricted solutions with no side flow in the *z* direction, solutions are only possible through extended iterative computer solutions. The order of magnitude of the solutions, however, can often be checked by reference to results for incompressible lubricants, especially for cases of low Λ(Ref 9).

Tilting-Pad Bearings

Some marked differences in behavior exist between tilting-pad gas-lubricated bearings (Fig. 2) and liquid-lubricated bearings. The expected pressure is somewhat modified, as shown comparatively in Fig. 7. The peak pressure is shifted toward the trailing edge of the shoe. This means that the pivot position for a gas-lubricated bearing is not the same as for a liquid-lubricated bearing. Figure 8 shows quantitative values for a tilting-pad gas-lubricated bearing for various values of Λ(Eq 2). In this case, $\Lambda$was reduced by lowering the ambient pressure.

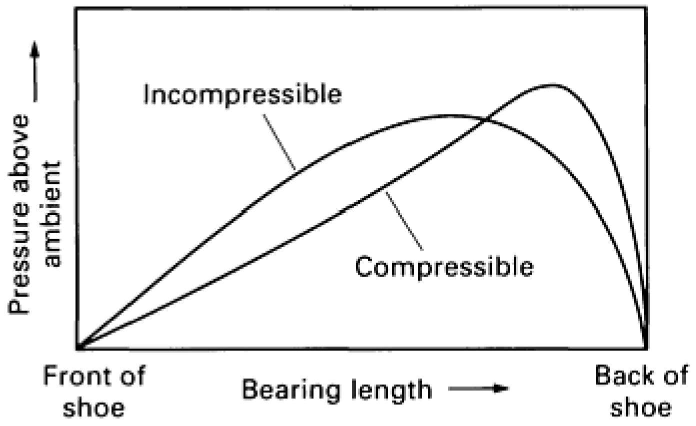


Fig. 7 Comparative pressure distribution for compressible and incompressible lubrication. Source: Ref 6

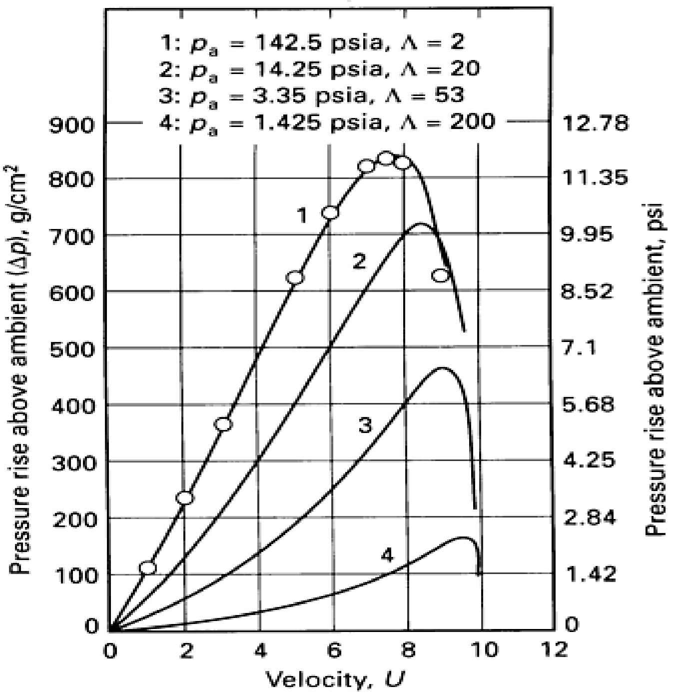


Fig. 8 Gas pressure distribution in flat tilting-pad bearing. Note open circles (data points) for theoretical liquid case. Source: Ref 7

Notice that for high values of Λ, the center of pressure moves toward the trailing edge or the back of the pad. For each designated operating condition, the pivot should be located at the center of pressure. Thus, if the ambient pressure, the speed, or the load is subject to change, a fixed pivot position will be optimum for only one set of operating conditions. This is a disadvantage.

In contrast, with incompressible lubricants, the tilting pad will react to a change in load or speed so that the ratio of inlet-to-outlet film thickness remains constant, even with the minimum film thickness getting smaller with increasing load. Thus, with liquid lubricants, once the pivot position is chosen, it will remain optimized for all changes of load, speed, or viscosity.

With gas-lubricated tilting-pad bearings, yet another complication arises. As the load increases, the minimum film thickness decreases, but since the pivot location is fixed, the pad angle of inclination begins to get smaller and eventually becomes parallel to the runner. An unstable condition will result, and the bearing will be unable to support the load and will collapse completely.

Collapse of the gas film will not occur for a convex curved surface on a tilting pad (Fig. 9). Because of the geometry, neglecting the influence of surface roughness asperities, a wedge must always be developed no matter how thin the film. An experimental study by Brunner *et al.* (Ref 12) shows that for a pad 19 mm ($\frac{3}{4}$ in.) wide by about 13 mm ($\frac{1}{2}$ in.) long in the direction of motion, even a crown height, δ, of as little as 0.25 μm will prevent collapse of the film and produce a load-carrying characteristic that continues to rise as the film reduces. Experience has shown that crown heights for optimum load-carrying capacity are approximately equal to the minimum film thickness, or about 7.5 to 10 μm.

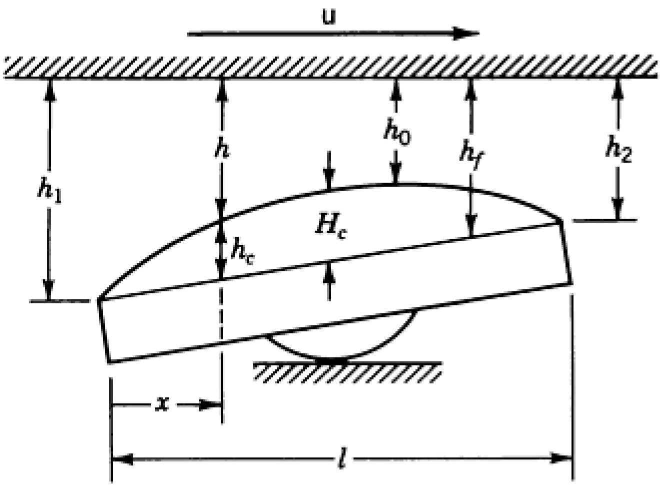


Fig. 9 Nomenclature for crowned tilting-pad bearing. Source: Ref 11

A thorough analysis of optimum crown heights and pivot positions can be found in Ref 4. Experimental verification of the theory has been especially good. Figure 10 shows values of the dimensionless load-carrying function as a function of the compressibility number for shoes having cylindrical crowns. The pivot position is not identified, but a value of about 0.6 is typical. That means, in Fig. 2, *c* is 0.6 of the length in the direction of motion and *a* is 0.4 of the length. The pivot is always between the midpoint and the trailing edge.

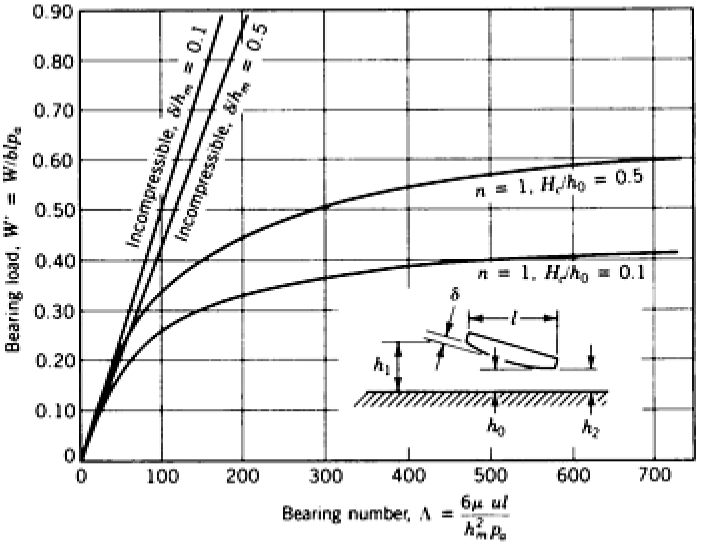


Fig. 10 Effect of bearing number on isothermal load for cylindrically crowned slider bearing, $b/l = 1$, $h_1/h_m = 2$, and $H_c/h_0 = 0.1$ and 0.5. Source: Ref 4

Effect of Mean Free Path

When gas films become very thin, as for example, when operating in a partial vacuum, the gas may no longer be considered as a continuous fluid, having the bulk viscosity of that fluid. Effects take place that can no longer be explained by continuous theory. Slip occurs at the boundary between the bearing surface and the gas and must be introduced into the calculation.

The measure of this slip condition is determined by the Knudsen number (K_n). This is the ratio of the mean free path of the gas (λ) to the film thickness in the bearing (h). Thus, $K_n = \lambda/h$. When K_n is less than 0.01, the gas flow can be treated as a continuum and slip flow ignored. When K_n is in the range of 0.01 to 1.5, slip flow increases and should be considered. When K_n is greater than 1.5, slip flow becomes significant and fully developed molecular flow results.

Typical values of mean free path at atmospheric conditions are:

Gas	Typical mean free path value	
	in. × 10^{-6}	**μm**
Hydrogen	4.43	0.1125
Helium	7.32	0.186
Air	2.52	0.064
Neon	5.20	0.132

Thus, with an air-film thickness of 2.5 μm (0.0001 in.), the value of K_n would be $2.52 \times 10^{-6}/100 \times 10^{-6}$, or 0.025. The slip effect would be small but not negligible.

Many tilting-pad-type thrust bearings with air lubrication are used in magnetic storage devices for computers, where they "fly" at film thicknesses of about 0.5 μm (20 μin.). At these very small film thicknesses, surface roughness can become a significant factor because no surfaces are perfectly smooth. This effect can be analyzed and included in predicting load-carrying capacity.

Rayleigh Step Bearings

This type of bearing is shown in Fig. 3. Analytical results for a square-step bearing (width equals length) are shown in Fig. 11 for lubrication with gas. A dimensionless load function *W*' is plotted against the compressibility number established for step bearings (see Eq 3).

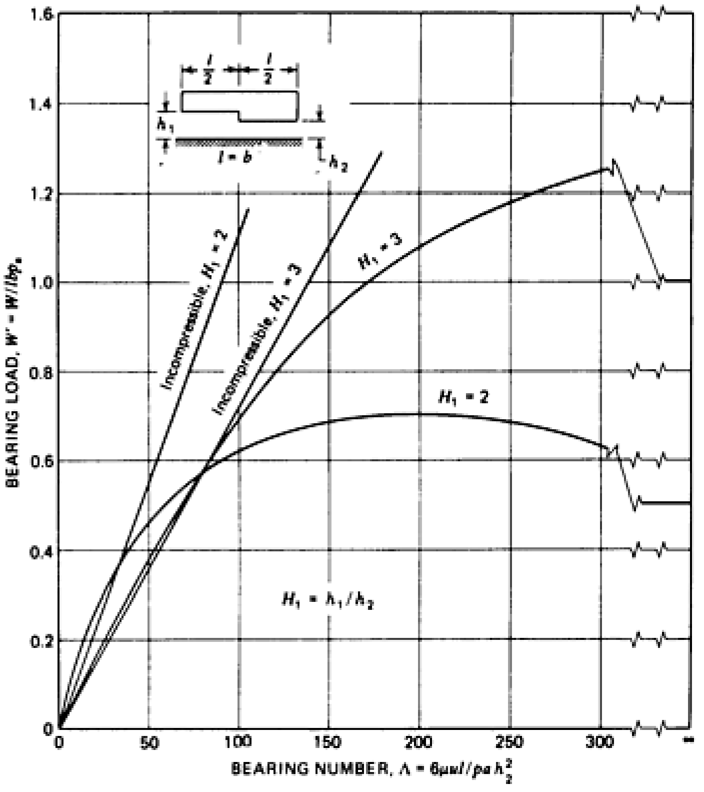


Fig. 11 Effect of bearing number on isothermal load for square-step slider bearings with film thickness ratios h_1/h_2 = 2 and 3. Source: Ref 4

Journal Bearings

This was the geometry first used to demonstrate true load-bearing ability by Kingsbury in 1897 (Ref 13). The same concept was used by Brubach (Ref 14) in utilizing inexpensive hypodermic syringes as air bearings. These are precision-made plungers in closely fitting barrels that perform remarkably well as hydrodynamic journal bearings.

Brubach has used these syringes as bearings for small centrifugal blowers running at 6000 rpm (Fig. 12). Perhaps his most important application has been in the measurement of gas leaks at very low pressures. Figure 13 shows the apparatus used for this purpose. The plunger of the syringe is caused to rotate by air impinging on the Scotch tape vanes. Due to this rotation, the plunger acts as an air bearing and is separated from the barrel by a hydrodynamic air film. Because the surfaces are separated from the barrel by a thin air film, the force required to move the plunger along the axis of the barrel actually approaches zero.

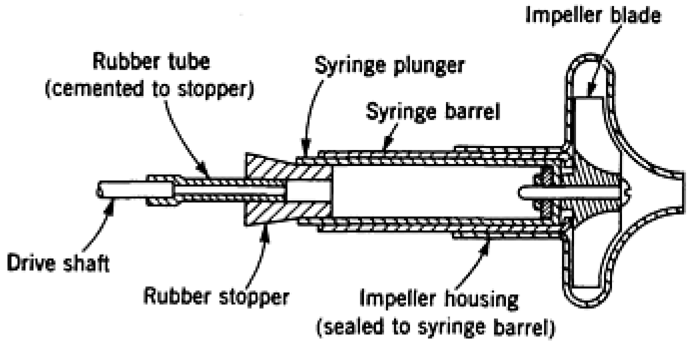


Fig. 12 Centrifugal blower using hypodermic syringe as air-lubricant bearing. Source: Ref 14

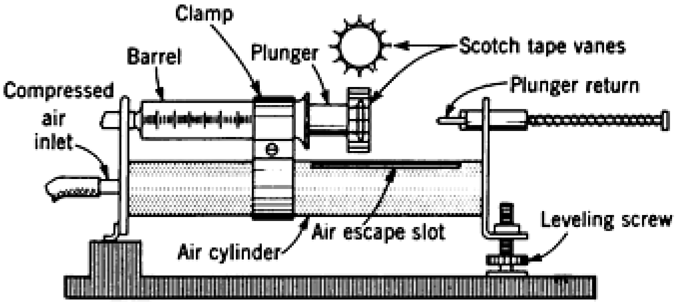


Fig. 13 Sensitive pressure-measuring apparatus using hypodermic syringe as air-lubricated bearing. Source: Ref 14

As an illustration of the sensitivity of this apparatus, it was connected to a tank of 100 L (26 gal) capacity. Then 0.1 mL (0.06 in.^{3}) of air was introduced into the tank. Immediately the plunger moved a corresponding 0.1 mL (0.06 in.^{3}). The theoretical pressure increase resulting from the injection of 0.1 mL (0.06 in.^{3}) into 100 L (26 gal) is about 0.01 mm (0.0004 in.) of water, and yet the low friction of the rotating plunger permitted it to respond instantly to this exceedingly small pressure differential.

Journal bearings are also influenced by compressibility of the gaseous lubricant. The value of Λ was given in Eq 1. Ford, Harris, and Pantall (Ref 7) show experimental correlation of the load-carrying parameter W' plotted against Λ (Fig. 14). For values of Λ greater than 1, the relationship is no longer linear. The deviation from a straight line indicates the point at which the compressibility effects make their influence felt.

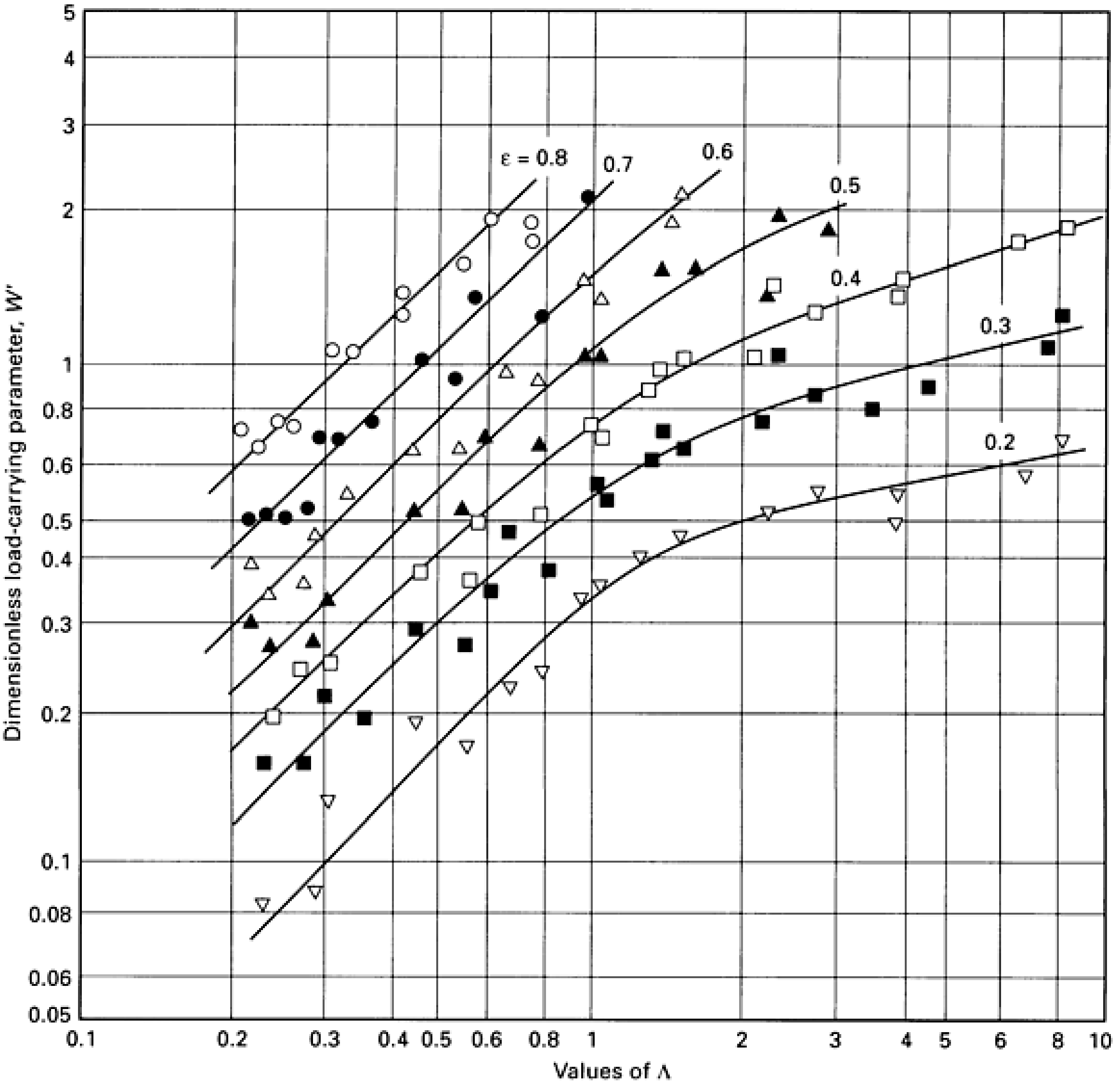


Fig. 14 Load parameter *W*' evaluated in terms of $\Lambda$and ϵ. Source: Ref 7

Another important variable in journal bearing analysis, especially stability analysis, is the attitude angle. In Fig. 1 this angle is shown as the angle, ϕ, between the line of load action, *W*, and the line of centers of the bearing, *O*-*O*'. The attitude angle will vary with load, speed, and other operating conditions, and also will depend on the geometrical configuration of the bearing itself. In general, if the attitude angle is small, the bearing is more stable against self-excited whirl. If the attitude is large, the bearing will be less stable against self-excited whirl.

For load-carrying capacity, the accumulated results of Gross (Ref 4), Elrod and Malanoski (Ref 15, 16), Elrod and Burgdorfer (Ref 17), and Raimondi (Ref 18) can be used as a general design approach for complete (360°) journal bearings. Figures 15, 16, and 17 show typical results from these analytical evaluations.

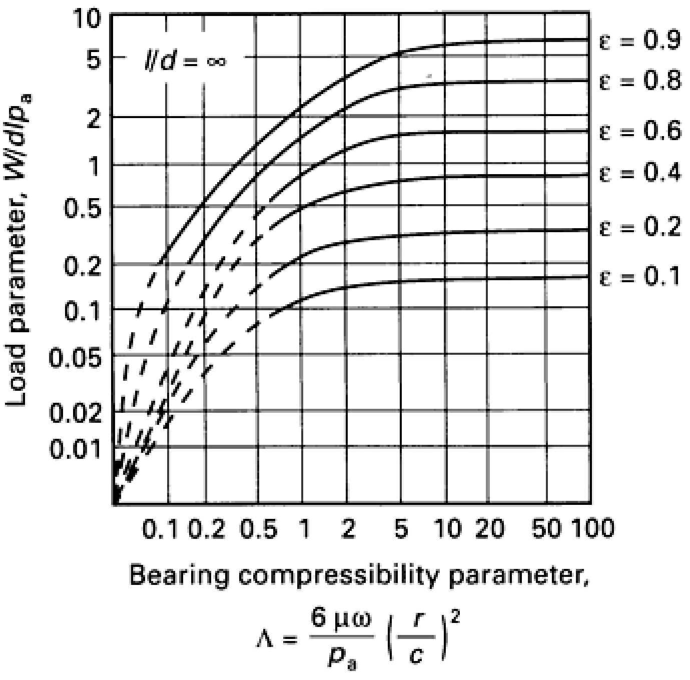


Fig. 15 Theoretical load-carrying parameter versus compressibility number for full journal bearing with *l*/*d* = ∞

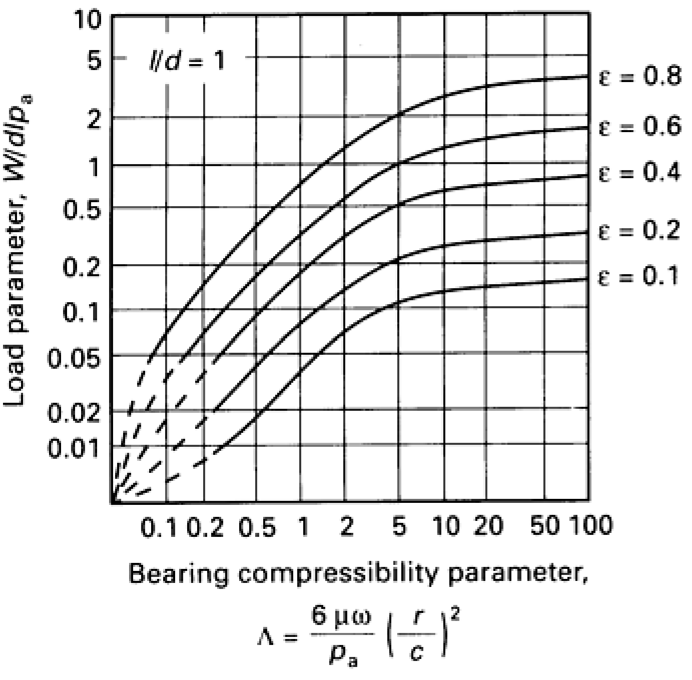


Fig. 16 Theoretical load-carrying parameter versus compressibility number for full journal bearing with *l*/*d* = 1

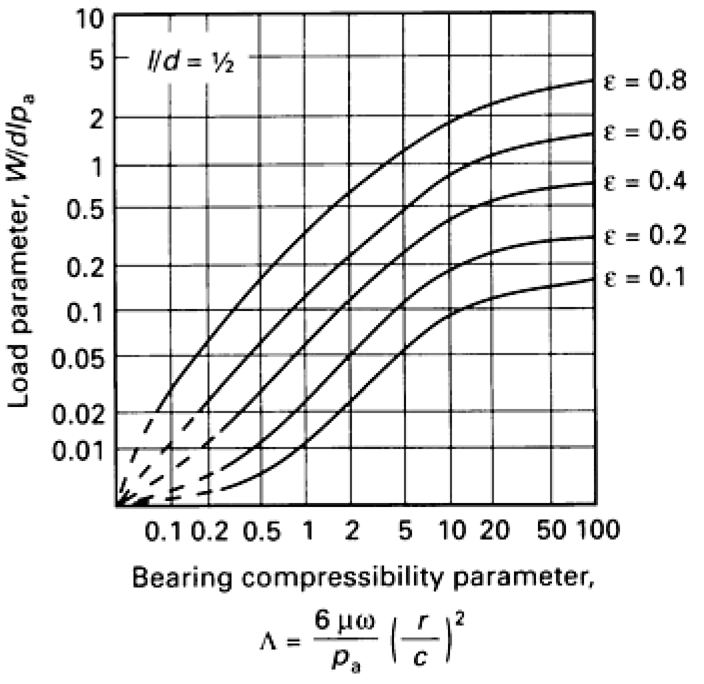


Fig. 17 Theoretical load-carrying parameter versus compressibility number for full journal bearing with *l/d* = $\frac{1}{2}$

Friction in Gas-Lubricated Journal Bearings

Of interest in some design applications is the power loss in the bearing due to friction resulting from viscous drag effects of the gas. Raimondi (Ref 18) has graphs of friction factors applicable to four *l*/*d* ratios, namely ∞, 2, 1, and $\frac{1}{2}$ for values of ϵ.

Specifically, the dimensionless friction factor for a plain journal bearing of ∞ length is defined as:

$$C_f^{\infty} = \frac{M_J}{2\pi\mu U r^2 l/c} \quad \text{(Eq 5)}$$

where,in English units, M_J is the frictional moment on the journal (in. · lbf). Figure 18 shows the friction factor plotted against the compressibility parameter.

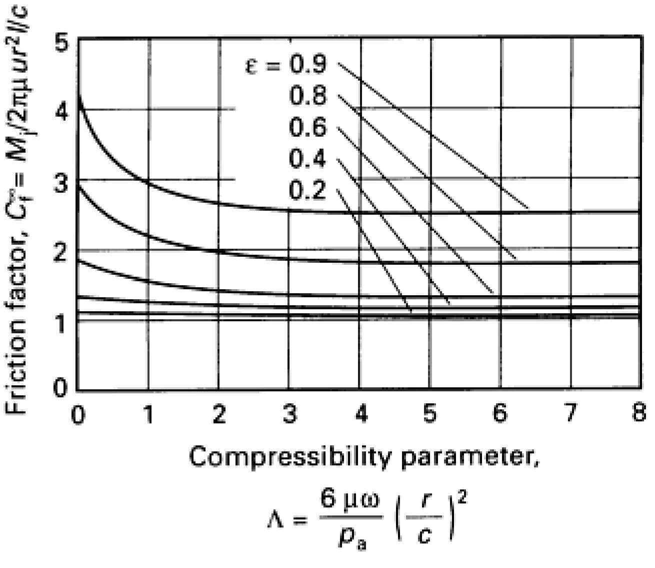


Fig. 18 Friction factor for full bearing of infinite length. Source: Ref 16

It can be seen from Fig. 18 that the friction loss is quite independent of Λ up to values of ϵ about 0.5. This means that for many bearings with low ϵ, values of friction can be reasonably estimated by using only simple Newtonian viscous shear. That would be shown by:

$$M_J = \mu A r \frac{v}{h} \quad \textbf{(Eq 6)}$$

where *A* is the swept area, which equals 2 πrl; *v* is the linear surface velocity; and *h* is the radial clearance, which equals *c*.

Synchronous Whirl

Synchronous whirl is a term used to describe a vibration condition in a fluid-film bearing where the mass of the rotor, combined with the spring stiffness of the bearing fluid film, can result in a natural frequency of vibration that is troublesome. If this natural frequency is the same as the rotational speed of the shaft, very large motions may be produced, resulting in possible bearing failure. If this natural frequency is below the operating speed, then it is necessary to pass through this resonant condition when approaching bearing speed.

Synchronous whirl can take the form of a translatory whirl, where the center line of the shaft describes the locus of a cylinder, or synchronous whirl can be conical. The conical mode form would resemble a bow tie, where there is no displacement at the center, only at the ends of the rotor. The displacement at the ends of the rotor would be out phase with each other.

If the bearing housing itself is elastically supported, it may also be capable of vibrating in either translatory or conical modes of vibration. These phenomena are analyzed in considerable detail in Ref 9.

A generally satisfactory approach is to estimate the quasi-static stiffness of the bearing film and then, in conjunction with the mass of the rotor, evaluate the natural frequency of vibration. It is a method that has proven its value over dozens of applications; Ref 9 provides examples of analyses that demonstrate the procedure.

Half-Frequency Whirl

One of the most serious forms of instability encountered in journal bearing operation is known as half-frequency whirl. This phenomenon is one of self-excited vibration and is characterized by having the center of the shaft orbit around the

center of the bearing (within the clearance circle) at a frequency of one-half of the spinning or rotational speed of the shaft (the speed may be little less than one-half in some applications).

Under these conditions, Hagg (Ref 19) has shown that the capacity of the bearing to support radial loads falls to zero. The shaft system may be stable as the speed is increased until this threshold is reached. Crossing this threshold with further increase in speed will bring the system into a region of instability. Unlike an ordinary critical speed, the shaft cannot pass through this one and attain a region of stability on the other side at a higher speed as with synchronous resonant whirl.

A typical instance would be a bearing running smoothly without any difficulty at 40,000 rpm, but seizing and falling completely at 43,000 rpm. Failure in most instances is instantaneous and complete as the amplitude of vibration becomes equal to the radial clearance in the bearing. This is especially true of rigid gas-lubricated bearings because of their inherent low damping.

This type of instability is not specific to gas bearings, but is a serious and continuing problem with high-speed, lightly loaded, rigid, liquid-lubricated bearings as well (Ref 9).

The compressibility of the gas has significant impact on the threshold of instability in gas-lubricated bearings, and must of course be included in the analysis. One of the more convenient procedures is due to Castelli and Elrod (Ref 20) and deals with translatory half-frequency shaft whirl in infinitely long, cylindrical bearings.

Although the procedure applies particularly to bearings of infinite length, it appears to have adequate accuracy in predicting the stability threshold for shaft in very long bearings--for example, where *l*/*d* values are 2.5, 3, or greater.

It also appears to be conservative for all 360° bearings of finite length, provided that the eccentricity ratio for static loads, ϵ_0, is calculated for an equivalent bearing of infinite length. This can be done using Fig. 15, with a unit load on the bearing (lb/in.) corresponding to that of the finite length bearing of interest.

The procedure is to use Fig. 19 from Ref 20, showing the relationship between ω^*_1, and $\Lambda$for several values of ϵ_0. These curves represent the threshold of instability for the bearing of interest. The value of ϵ_0 is computed on a static load basis:

$$\omega^*_1 = \omega \sqrt{\frac{cM_1}{w}} \qquad \textbf{(Eq 7)}$$

where $\omega$is the shaft speed (rad/s), *c* is the radial clearance (in.), M_1 is mass per unit length (lb · s^2/$in.^2$), and *w* is load per unit of length (lb/in.). Individual curves are drawn for ϵ_0, the value of eccentricity ratio for a bearing of infinite length but with the same unit loading (lb/in.) as the bearing being examined.

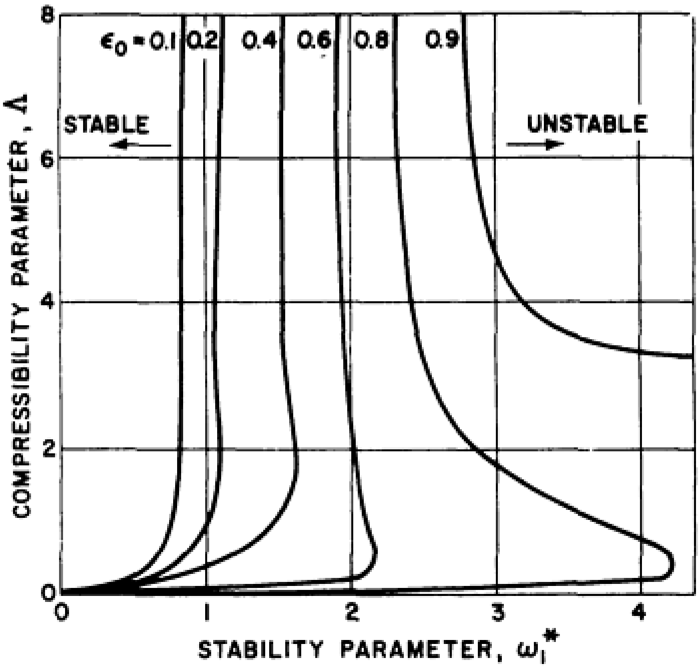


Fig. 19 Plot of half-frequency whirl threshold for infinite-length 360° journal bearing. Source: Ref 20

If the intersection of ω^*_1 and Λ falls to the left of the curve for the operating value of ϵ_0, the bearing should be stable. If it falls to the right, the bearing should be unstable.

Tilting-Pad Journal Bearings

One favored bearing design to achieve rotor stability is to use tilting-pad journal bearings. Considerable analytical and experimental work has been devoted to these bearings, both in the liquid- and gas-lubricated forms. These bearings also offer circumferential and axial misalignment capabilities.

Placing a resilient support (like a spring) under one pad of a three-pad bearing permits accommodation of centrifugal growth of the shaft at high speed or thermal or elastic dimensional changes in both shaft and bearing housing. When properly designed, this bearing is very forgiving. Many succesful applications have been made.

When using gas as the lubricant, some designs will include an auxiliary hydrostatic lift for starting (Fig. 20), which will reduce initial solid contact friction and protect the surfaces. The complete design is quite complex and an area of great specialization. Design guides for gas-lubricated journal and thrust bearings can be found in Ref 21, 22, 23, 24, and 25.

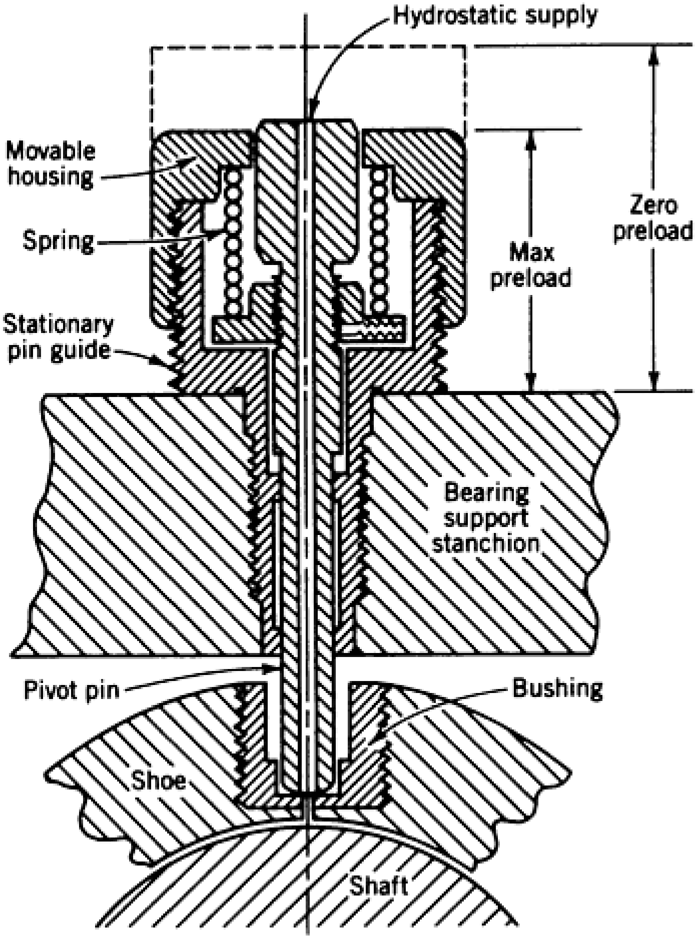


Fig. 20 Cross-sectional view of a spring-mounted pivot assembly. Source: Ref 21

A few fundamental concepts are critical in the design of a tilting-pad journal bearing. Of primary importance are yaw stability, pivot circle clearance, and pivot design. Other design features--such as film stiffness of combined bearing, minimization of bearing friction, the effect of clearance on minimum film thickness, shoe pitching frequency, and rotor critical speeds--are examined in the referenced design manual (Ref 21).

Yaw Stability. A pivoted shoe (or pad) has three degrees of freedom. It can pitch and roll like a ship, and it can yaw. To yaw is to act like a boat rotating about a vertical axis through its center of gravity. Any such rotation will lead to film rupture at the outer corners of th pad, with resulting solid contact. To provide for yaw stiffness in the bearing pad, the angle of wrap of the shoes should be no less than 90° of circumferential arc. With shorter arcs, the pad rapidly loses its ability to provide a resisting moment.

Pivot Circle Clearance. One of the crucial parameters governing the operation of the bearing is the setting of the pivot circle clearance (C' in Fig. 21). If the shaft always remained in the central concentric position, then C' could equal the actual clearance in the bearing, designated by C. That is to say, the radius of curvature of the shoe equals the radius of curvature of the shaft, R, plus the clearance C. The radius of curvature of the shoe is thus greater than the radius of curvature of the shaft. This produces the same phenomenon as having a crowned tilting pad. The load-carrying film cannot collapse even if the film becomes very thin.

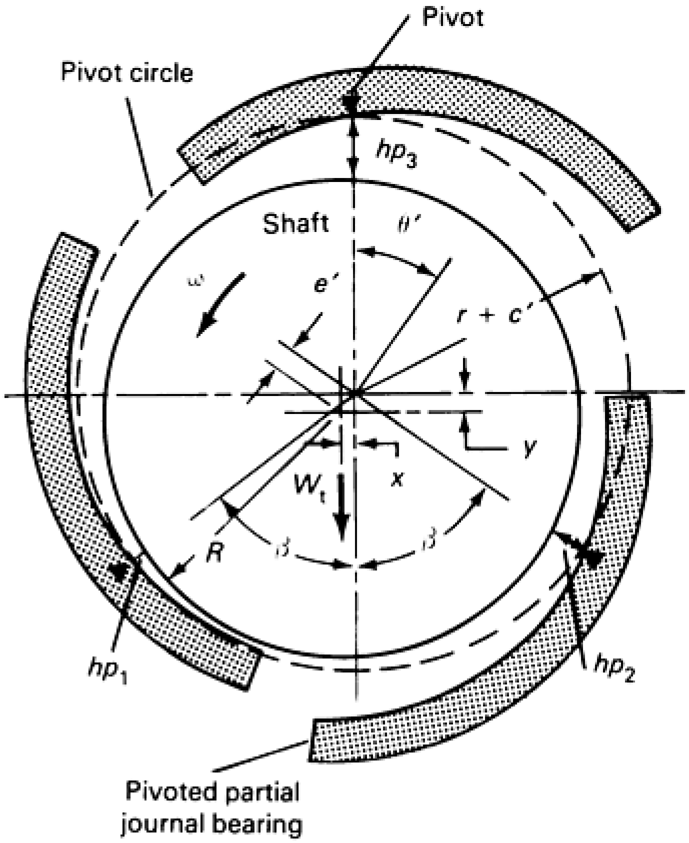


Fig. 21 Pivoted-pad journal bearing with three shoes. Source: Ref 21

However, with the journal bearing under load, and with low speeds, the eccentricity of the shaft can become great enough so that it moves away from the top (unloaded) tilting pad, so that the pad will lose its converging wedge and become unstable. To avoid this possibility, the shaft is preloaded by squeezing the pads radially inward, making *C*′ less than *C*. The measure of this is the ratio *C*′/*C*, which should be less than 1. This is a design parameter and depends on the expected range of eccentricity ratios to be encountered.

Pivot Design. The third important concept is the design of the actual pivot. A ball and socket joint will not be satisfactory because of the frictional resistance to sliding. As the pad assumes its equilibrium position, its righting capacity goes to zero. Any residual friction in the pivot will prevent the pad from ever reaching its equilibrium design inclination. Figure 22 shows a recommended configuration utilizing rolling friction rather than sliding friction. Incidentally, the surfaces should be smooth and very hard to prevent fretting corrosion.

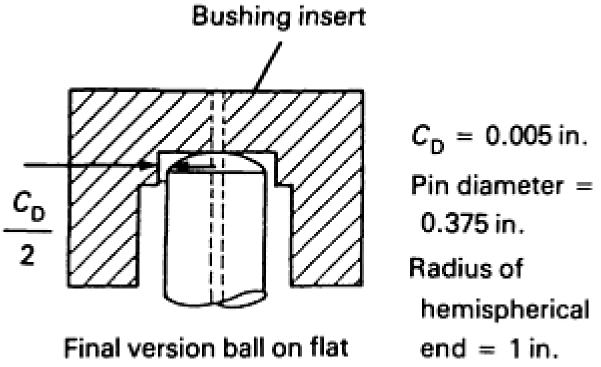


Fig. 22 Bushing insert for pivoted-pad hydrostatic journal bearing. Source: Ref 21

Three-Sector Journal Bearing

This journal bearing, with three equally spaced axial grooves, has shown some degree of stability against self-excited whirl. Castelli and Pirvics (Ref 26) have presented comprehensive numerically computed performance characteristics for three- and four-axial grooved gas-lubricated journal bearings. Angular extent of each groove is considered to be 5°. Thus for a three-sector bearing, the arc length of a sector would be 120 - 5 = 115°.

Figure 23 lists results in terms of the dimensionless load capacity parameter $W' = W/p_a\ rl$ plotted against the bearing compressibility number Λ. Note that the value of Λ used here is actually the same as in Eq 1. The applied load is directed toward the center of one sector. Evaluating W' and Λ will determine the eccentricity ratio and thus the minimum film thickness.

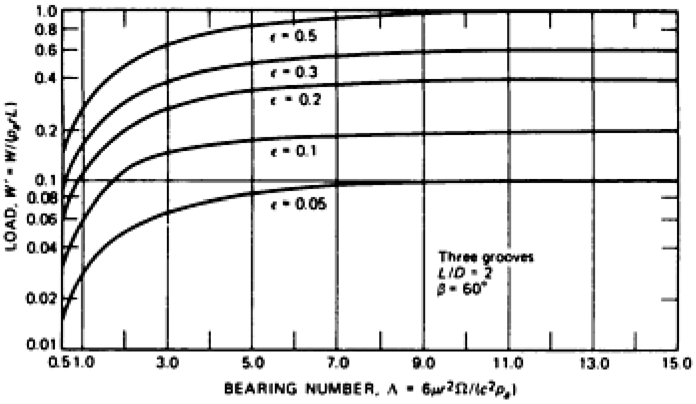


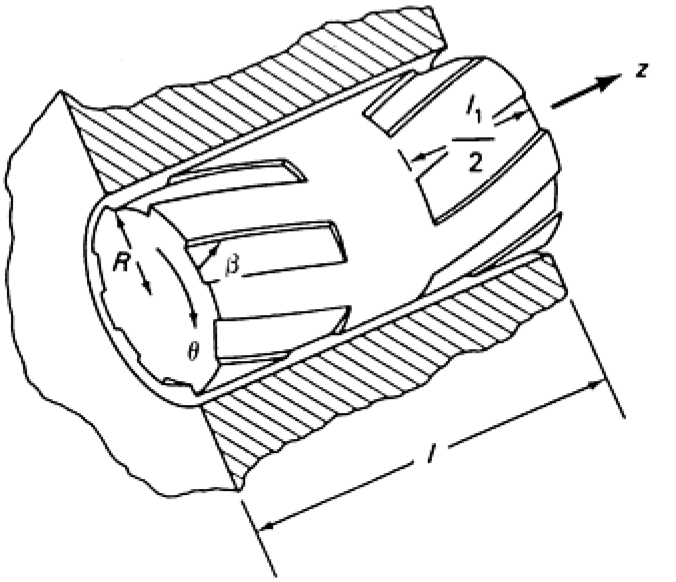
Fig. 23 Load function *W'* versus Λ for three-groove journal bearing. Source: Ref 26

Information on many other sizes, load directions, and attitude angles can be found in Ref 26 and 27.

Helical-Grooved Journal Bearings

The helical grooving in this type of journal bearing enhances stability by reducing the attitude angle below that obtained from a plain cylindrical journal bearing. These bearings are known for their stability and are often used as a possible substitute for tilting-pad journal bearings.

Castelli and Vohr (Ref 28) solved the appropriate equations numerically for load capacity and attitude angle for the case of l/d = 1.0, with various values of Λ. Figure 24 lists the geometric parameters for the spiral-grooved bearing as used in Ref 28. Figure 25 lists the results showing the load parameter $W' = Wp_a$ as a function of Λ with the eccentricity ratio ϵ as the third variable. Malanoski (Ref 29) shows good comparison between the theoretical predictions of Castelli and Vohr and his own measured results for helical-grooved journal bearings.



$$\alpha = \frac{a_g}{a_g + a_r} \qquad \bar{z} = \frac{l_1}{l} \qquad \Gamma = \frac{h_g}{h_r}$$

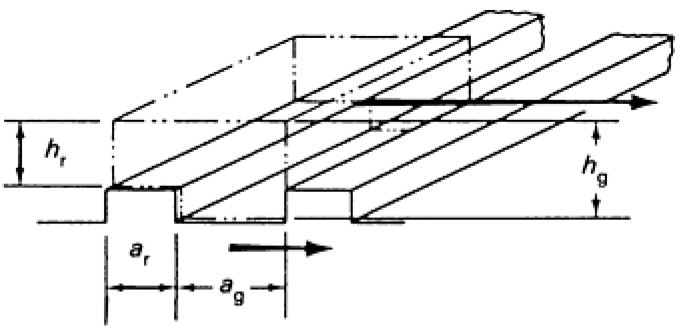


Fig. 24 Geometry of spiral-groove bearing, using notation of Castelli and Vohr (Ref 28)

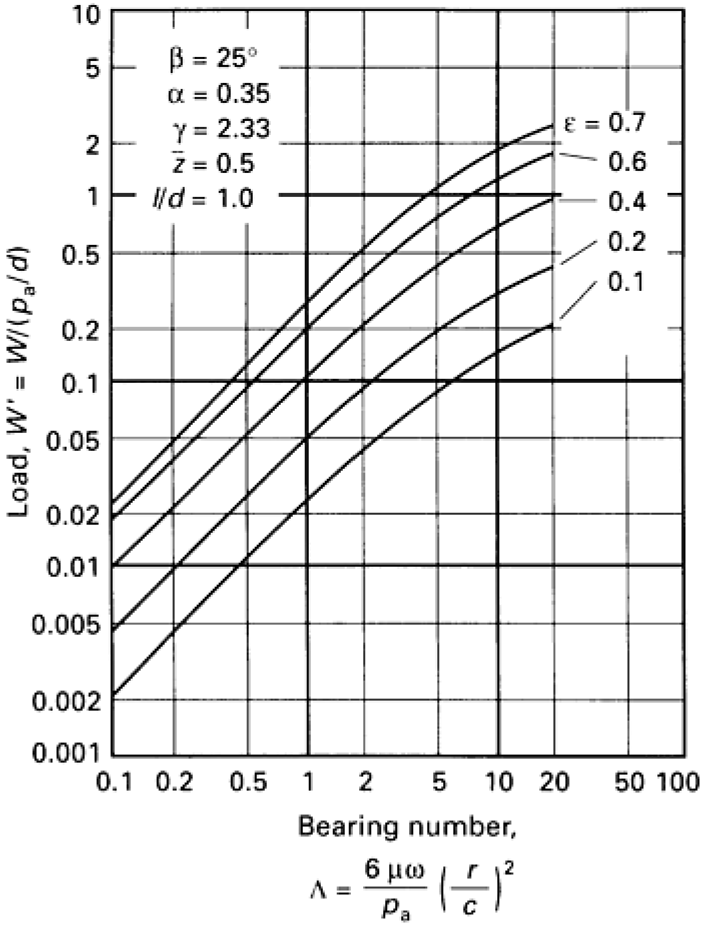


Fig. 25 Load capacity of spiral-groove journal bearing as a function of bearing number and eccentricity ratio. Source: Ref 28

Hydrostatic Gas-Lubricated Bearings

The many advantages of externally pressurized bearings are well known (Ref 9). With gas they have the added benefit of extreme cleanliness, and the use of gas enables them to operate over a wide range of temperatures.

However, analytical and design complications arise because of the compressibility of the gas. At low supply pressures (gage) equal to or less than ambient pressure (absolute), the system can be very simple; for example, supply pressure 70 kPa (10 psig) with an ambient pressure of 100 kPa (14.7 psia).

With high feed pressure, the gas flow in the entrance section is extremely complicated and may involve choked flow, shock waves, vortex formation, and boundary layer growth. Many comprehensive studies have been made of supersonic pressure depression in the feeding region of externally pressurized bearings (Ref 30, 31, 32, 33, 34).

These bearings do not involve a constant volume of flow as is the case with many liquid-lubricated hydrostatic bearings. Therefore, in order to achieve stiffness, they must have some kind of upstream restrictor in the feed line (Fig. 26). The flow restrictor can be an orifice or a capillary, and the bearing is then described as being restrictor compensated; the orifice restrictor area is equal to $\pi d_0^2/4$ (Fig. 26).

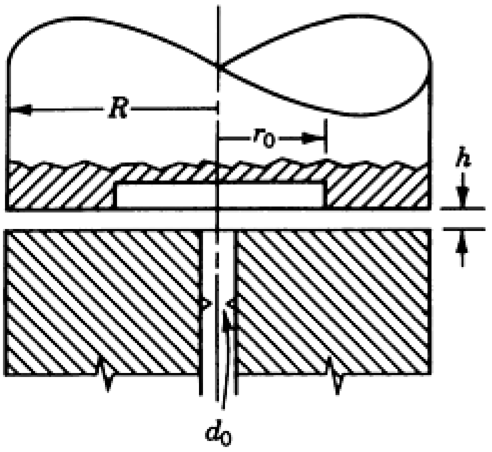


Fig. 26 Flow-restricted hydrostatic gas-lubricated bearing

Sometimes, the resistance to flow at the entrance to the film itself may dominate. In that case the bearing is identified as having inherent compensation. Many bearings are of this type. The inherent restrictor area would then be the circumferential annulus $\pi r_0 h$ at the entrance to the film (Fig. 26).

A typical pressure profile is shown in Fig. 27 for a simple circular thrust bearing with a central feed source. With a supply pressure of 480 kPa (70 psig) and a film thickness on the "sill" of *h* = 0.05 mm (0.002 in.), the effect of sonic velocity is seen. The minimum pressure measured on the sill (p_s) is 76 kPa (11.01 psia), indicating a partial vacuum. The maximum recovery pressure on the sill is 101 kPa (14.6 psig).

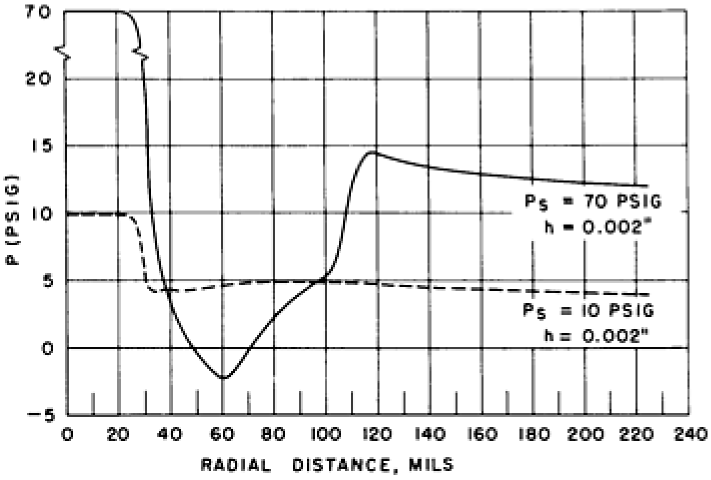


Fig. 27 Experimental pressure profiles for simple area 2 πr_0 *h* externally pressurized thrust bearing

However, notice in Fig. 27 that the radial distances are measured in mils (0.001 in. = 0.025 mm), so that all of this "micro-aerodynamic" activity has taken place within a radius of 120 mils, or about 3.2 mm. The remainder of the bearing area can thus be treated as laminar isothermal gas flow and analyzed accordingly.

The dashed line in Fig. 27 shows the pressure profile when the bearing feed pressure was 70 kPa (10 psig). There is no sonic flow. Allowing for the pressure drop in the restrictor, either external or inherent, the bearing characteristics can be obtained in a relatively simple manner.

Figure 28 is a sample of the excellent program conducted by Laub (Ref 35) on thrust bearings and journal bearings intended for metrology applications. The pressure profiles show no indication of significant flow restriction and localized pressure loss because of bearing geometry and low pressure levels that were used. Flow evaluation, however, must include compressibility effects.

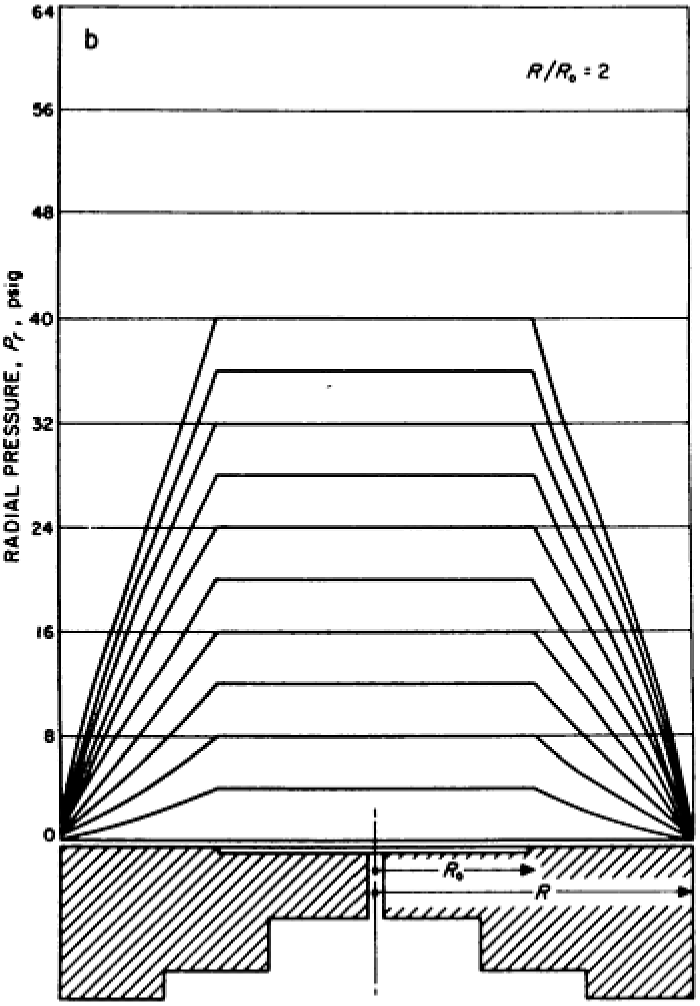


Fig. 28 Pressure profiles in gas-lubricated hydrostatic bearing: Source: Ref 35

As an example of the need to include the compressibility effect in flow prediction, consider the hydrostatic step bearing shown in Fig. 29. For liquids (incompressible), the supply pressure, P_0 and the ambient pressure, P_1, are gage pressures, R_0 is the radius of the recess, and R is the radius of the shaft. The derivation for the flow of lubricant in such a bearing is given in Ref 9 as:

$$Q = \frac{P_0 \pi h_0^3}{6\mu \ln (R/R_0)} \quad \textbf{(Eq 8)}$$

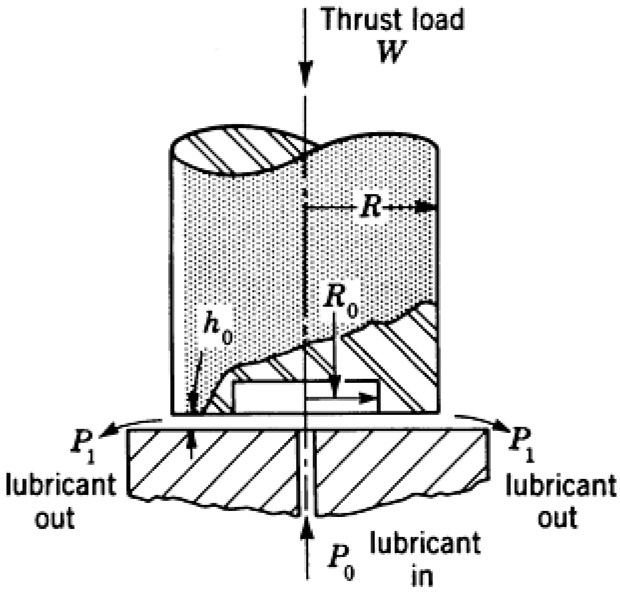


Fig. 29 Schematic diagram of a step bearing

However, when compressibility effects are included, the equation for flow volume becomes (Ref 9):

$$V_0 = \frac{\pi h_0^3}{6\mu \ln \dfrac{R}{R_0}} \left[\frac{P_0^2 - P_1^2}{2P_0} \right] \qquad \textbf{(Eq 9)}$$

where P_0 and P_1 are in this case absolute pressures, and V_0 is the flow volume at the supply pressure P_0.

Pneumatic Hammer

The most troublesome characteristic of the externally pressurized air bearing is instability. During test programs, the phenomenon of self-excited vibration is often encountered, characterized violent fluctuations of pressure in the recess and amplitudes of vibration many times greater than the gap width at the equilibrium point. This phenomenon is often called pneumatic hammer.

Licht, Fuller, and Sternlicht (Ref 36) used a simplified lumped-parameter analysis to examine this problem. The gas film density in an oscillating thrust bearing is time dependent, and, in general, the mass inflow does not equal the mass outflow. As a consequence of film compressibility, energy from the film may be periodically added to the system in phase with the motion so that instability develops. The vibration is independent of system resonances.

The general stability analysis reveals the following:

- For a constant supply pressure, p_s, stability is enhanced by increasing the recess pressure p_0--that is, minimizing the pressure drop through the supply restrictor so that p_s - p_0 is a minimum. This of course reduces the stiffness of the bearing
- A recess depth comparable to the film thickness would be the ideal
- Maximizing the size of the inlet supply orifice will increase stability because capillary restriction is more likely to be unstable than orifice restriction
- Incompressible films are always stable

Multiple-Pressure Sources

To avoid pneumatic instability, it is clear that the high-pressure recess should be very shallow. The limit would be a recess of zero depth or no recess at all. However, that would reduce the load capacity of the bearing. A frequently used alternative is a ring or other appropriate pattern of multiple supply orifices that acts to develop an equivalent high-pressure area with the same depth as the film itself. Figure 30 shows a simple thrust bearing modified in this manner.

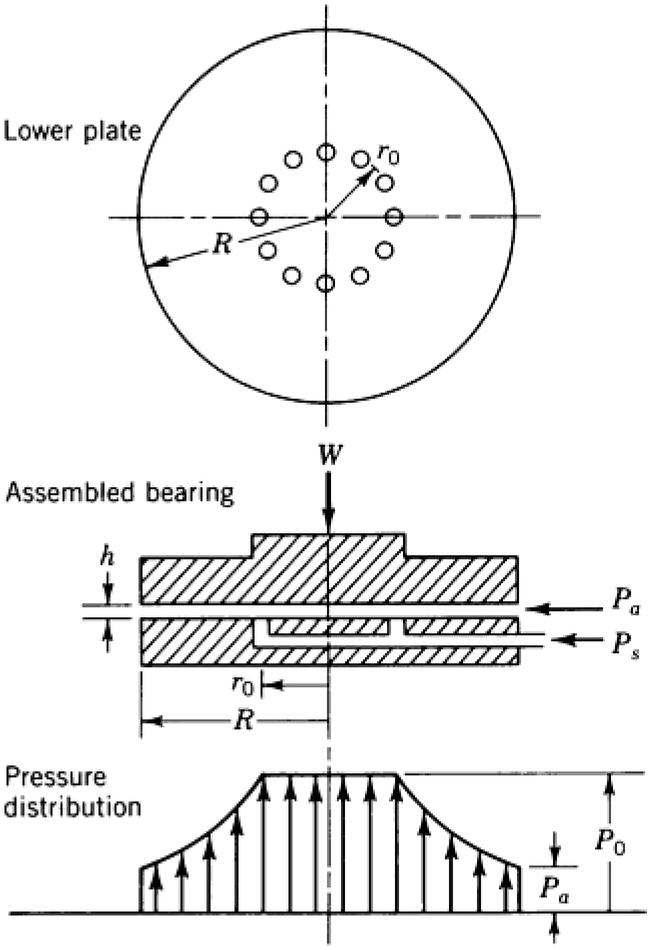


Fig. 30 Multiple-source feed for thrust bearing

In this bearing, the orifices are located on a circle of radius r_0. All orifices feed air to the interface at the same pressure P_0. Because there is no pressure gradient between the orifices, there is no flow between the orifices, and the entire circle of radius r_0 acts as a high-pressure recess. The same concept has also been successfully applied to journal bearings by Laub (Ref 35).

Porous Bearings

An alternate means for reducing the size and depth of a pressurized recess, other than using a finite number of multiple orifices as just described, is to feed the bearing through a section of porous material. Gas is admitted to the bearing interface through the pores of the material, resulting in a very large number of feeding restrictors in parallel. Again, the recess has been eliminated and stability enhanced. Many porous bearings have been made in both flat thrust and cylindrical journal bearing configurations. Sneck (Ref 37) provides an excellent survey of the many applications that have been made of this type of bearing, plus a very complete list of references.

Figure 31 is typical of a circular thrust bearing with pressurization through a porous annulus.Frequently, a porous carbon graphite is used so that antiscuff protection is provided by the material when in solid contact. A reasonable range of

permeability is available in these commercial products, and they have proven to be satisfactory. Clean air is essential to keep the pores from clogging with dust.

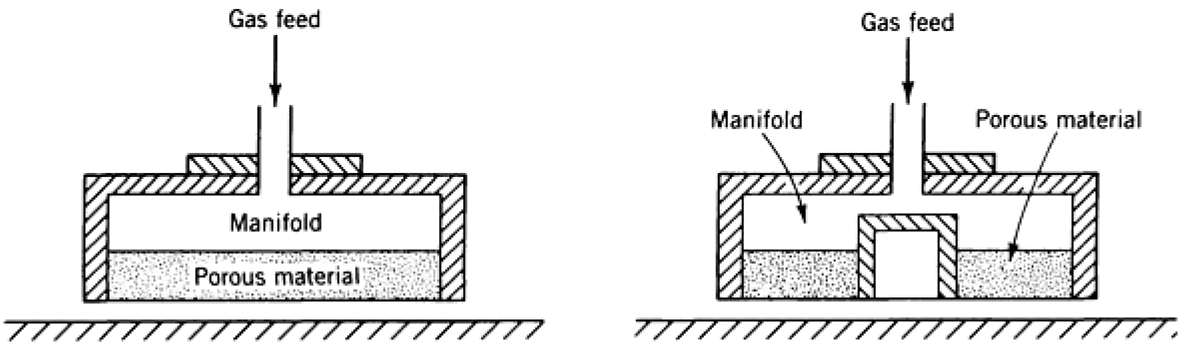


Fig. 31 Typical configurations for externally pressurized porous gas-lubricated thrust bearings. Source: Ref 9

Typically, pressurized porous bearings can be used anywhere classical orifice-compensated bearings are used. Design charts have been prepared by Gargiulo and Gilmour (Ref 38) to assist in a more exact analysis of these bearings.

However, it must be expected that the actual permeability of the material will be subject to some variation, even when cut from the same block. Vohr (Ref 4) discusses many additional design details involved in the use of porous materials in externally pressurized gas-lubricated bearings. These bearings can be extremely useful and are an attractive option for applications that call for a hydrostatic gas bearing.

Compliant-Surface Bearings

Compliant-surface bearings have been mentioned before in this article. They can use elastomers as the bearing material; in this form they have remarkable low-speed fluid-film capabilities. Foil bearings fall into this category as well. Compliant-surface bearings can be used as flexible membrane bearings.

The advantages of compliant-surface bearings include:

- Freedom from precision machining and maintenance of close tolerances
- Ability to accept misalignment
- Tolerance of dirt and particulates
- Accommodation of surface roughness with low surface speeds

The foil bearing (Fig. 32) is the most widely used form of compliant-surface bearing. It was first introduced, in simple form, by Blok and Van Rossum (Ref 39).

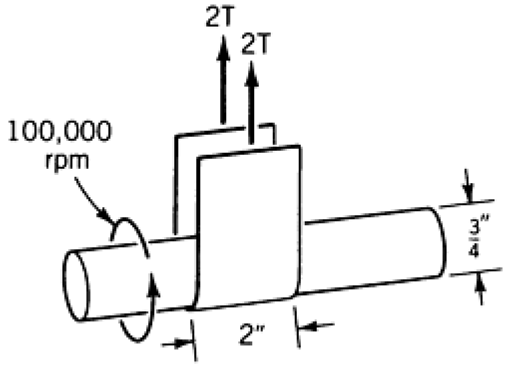


Fig. 32 Schematic diagram of a foil bearing. Source: Ref 27

The bearing can consist of a thin strip of flexible material (such as a plastic tape or thin metallic foil) partially wrapped around simple journal like the saddle belly band on a horse. As the journal spins, a reasonably large force can be supported by the self-acting hydrodynamic film in the contact area between the tape and the journal.

The simple foil bearing was further analyzed by Patel and Cameron (Ref 40) in 1957. It was utilized in the United States as an air bearing for applying load to a rotating shaft as early as 1956 by Fischer, Cherubim, and Fuller (Ref 27). By far, the greatest value of this simple concept is in tape transport for high-speed magnetic tape recorders. In this application, the journal is stationary and contains the recording head for the read-out components, while the tape glides past. Foil bearing analysis and design are reviewed extensively in Ref 4 and 41.

Developments of the original foil bearing concept of Blok and Van Rossum have now reached the stage of commercial application. Their advantages are many. If a metallic foil is used, the bearing can operate at high temperatures, especially when lubricated with air or some other gas. There is no problem with the possible loss of clearance due to differential thermal expansion between shaft and bearing, as is often the case with rigid surface units. The foil bearing establishes its own operating film thickness at all times.

It can also tolerate misalignment. In manufacturing, the foil bearing greatly reduces the need for holding expensive dimensional tolerances. An additional benefit is its stability in conjunction with high-speed rotor applications. The foil bearing is often used just for this reason, because it effectively reduces the possibility of self-excited fractional-frequency whirl (Ref 42).

Three distinct commercial varieties of foil bearings are available:

- Tension-dominated foil bearings (Fig. 33a)
- Bending-dominated segmented foil bearings (Fig. 33b)
- Bending-dominated continuous foil bearings (Fig. 33c)

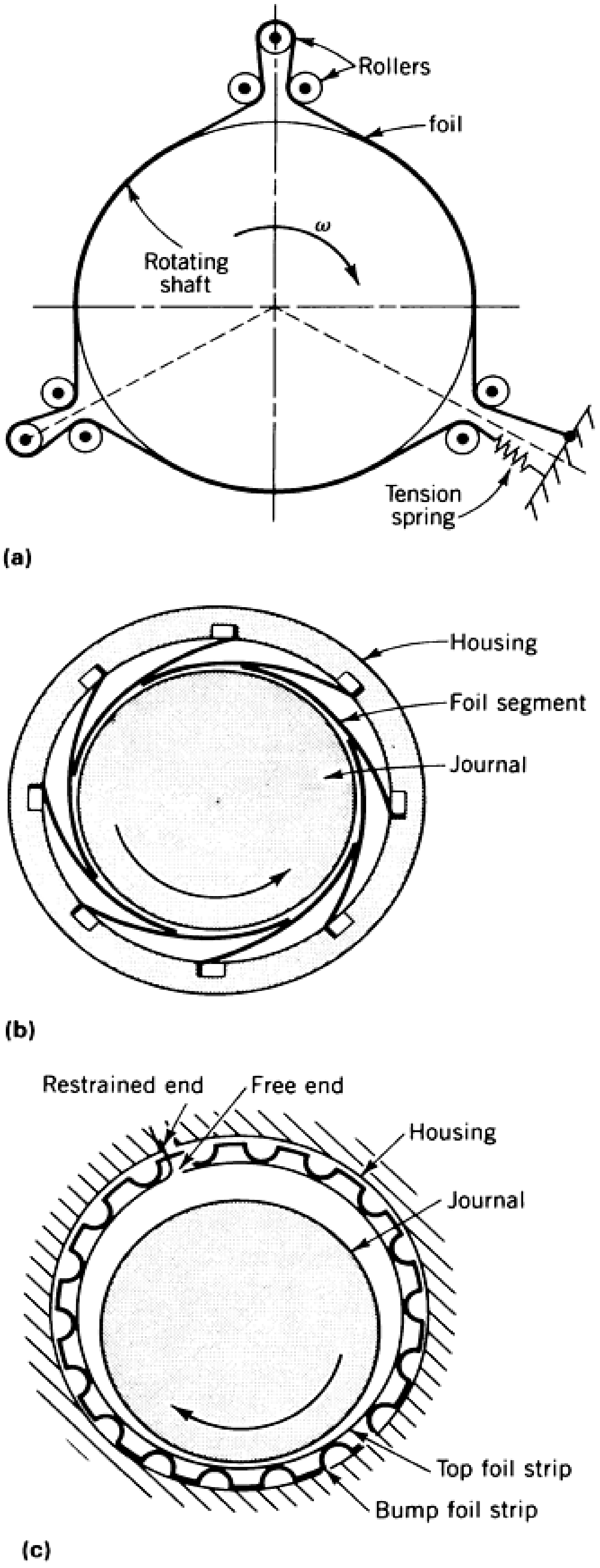


Fig. 33 Foil bearing designs. (a) Tension-dominated foil bearing. (b) Bending-dominated segmented foil bearing. Source: Ref 9. (c) Bending-dominated continuous foil bearing. Source: Ref 9

Tension-dominated foil bearings (Fig. 33a) have been used in data-processing equipment, where they provide support for magnetic tape traveling over guides, heads, and vacuum columns in tape transport. Recent developments have concentrated on adapting these bearings to the support of high-speed rotors for turbomachinery (Ref 43).

Bending-dominated segmented foil bearings of the type shown in Fig. 33(b) are widely used in both commercial and military aircraft applications. One example is the air-cycle refrigeration compressor used on many commercial transports for cabin cooling. Operating speeds can easily be 100,000 rpm, yet the compressor exhibits a high degree of reliability and freedom from rotor whirl instabilities. These bearings are being developed for cryogenic compressors at working temperatures near absolute zero.

A great deal of research and development work is being devoted to extending these bearings to high-temperature applications. With support from the Air Force Aero Propulsion Center and the Naval Air Propulsion Center, research programs are underway to develop a foil bearing capable of operating at high speed in a 650 °C (1200 °F) ambient. The potential application is for high-temperature gas turbines (Ref 44). Thrust bearing configurations are also used.

Bending Dominated Continuous Foil Bearings. This configuration has a single top foil strip that is restrained at one end and supported by a resilient corrugated foil strip called the bump foil strip (Fig. 33c). This piece elastically supports the top foil and controls the stiffness of the bearing. The deflection of the bump foil tolerates load fluctuations, and its elastic behavior provides resilience. Friction damping is also introduced with its beneficial stabilizing effects on the dynamics of the rotor.

Significant accomplishments have been recorded for bending-dominated continuous foil bearings, including:

- Three years of operation in automotive-type gas turbines at 60,000 rpm and 260 °C (500 °F)
- 120,000 rpm in product hardware
- Bearing shock resistance to 25 g at 100,000 rpm
- Tests at 650 °C (1200 °F)
- Unit bearing pressures for hydrodynamic air operation up to 345 kPa (50 psi), based on projected area under laboratory conditions

Modified Bending-Dominated Continuous Foil Bearings. This type of bearing is a modification of the design shown in Fig. 33(c). Although only limited applications have been made, it does appear to have excellent performance characteristics.

Design Analysis. There is no significant design analysis available, in the public domain, for the foil bearings described above. The design procedures are essentially proprietary. However, if a proposed application involves light starting loads ($\leq$7 to 14 kPa, or 1 to 2 psi) and when up to speed (>20,000 rpm) the unit loads do not exceed 70 to 100 kPa (10 to 15 psi), it appears that some type of foil bearing would be an excellent choice.

Pressurized-membrane bearings are bearings with a flexible membrane pressurized from the center through port holes (Fig. 34). At equilibrium conditions, the air escapes from under the inflated membrane through the minimum air gap. Levy and Coogan (Ref 45) first developed the analysis for such a device.

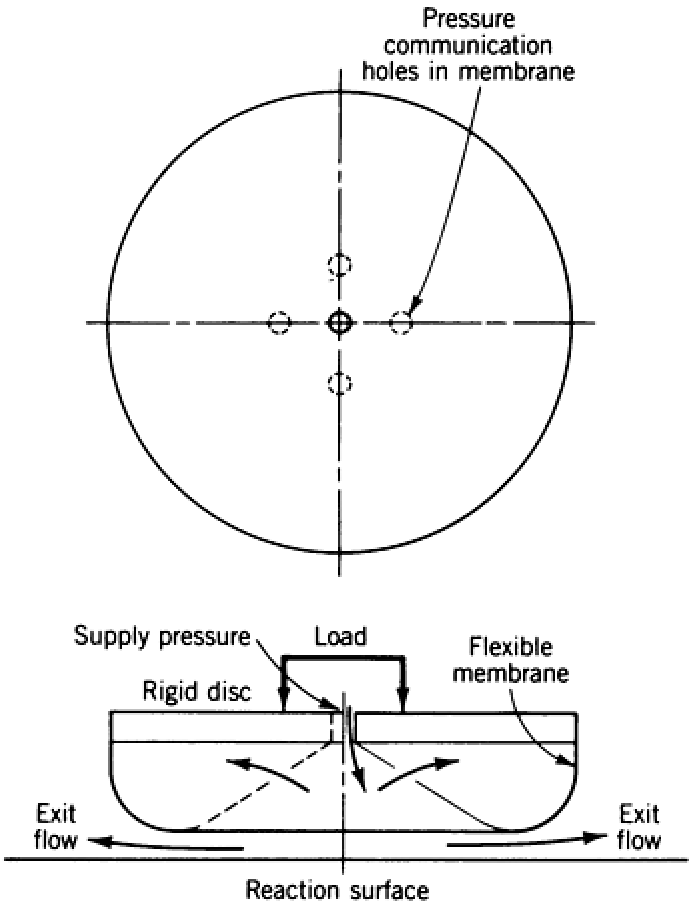


Fig. 34 Compliant-surface membrane bearing. Source: Ref 45

These bearings find wide use in areas where large loads must be lifted and transported over surfaces that may not be perfectly smooth, such as the floors in warehouses and factories. The flexibility of the membrane accommodates itself to the undulations that may exist, and motion proceeds smoothly with almost no friction. When the desired location or position is attained, the air supply is shut off and the object is set down. These devices are now available commercially.

A spectacular application of this type of hydrostatic membrane bearing is the Mile High Stadium in Denver, Colorado. The stadium was designed so that it would be suitable for both football and baseball. Accordingly, the entire side section of the grandstand, weighing 4080 Mg (4500 tons) and including 21,000 seats, is mounted at 46 points on water-lubricated hydrostatic rubber pads. Each pad is 1.2 m (4 ft) in diameter and made of fabric-reinforced synthetic rubber.

When pressurized, the entire section of the grandstand floats on a film of water and is moved back a distance of 44 m (145 ft) with very little effort. It is then set down in place, producing an arena suitable for playing baseball (Fig. 35). Although water was the pressurized lubricant of choice in this application, air could have been used.

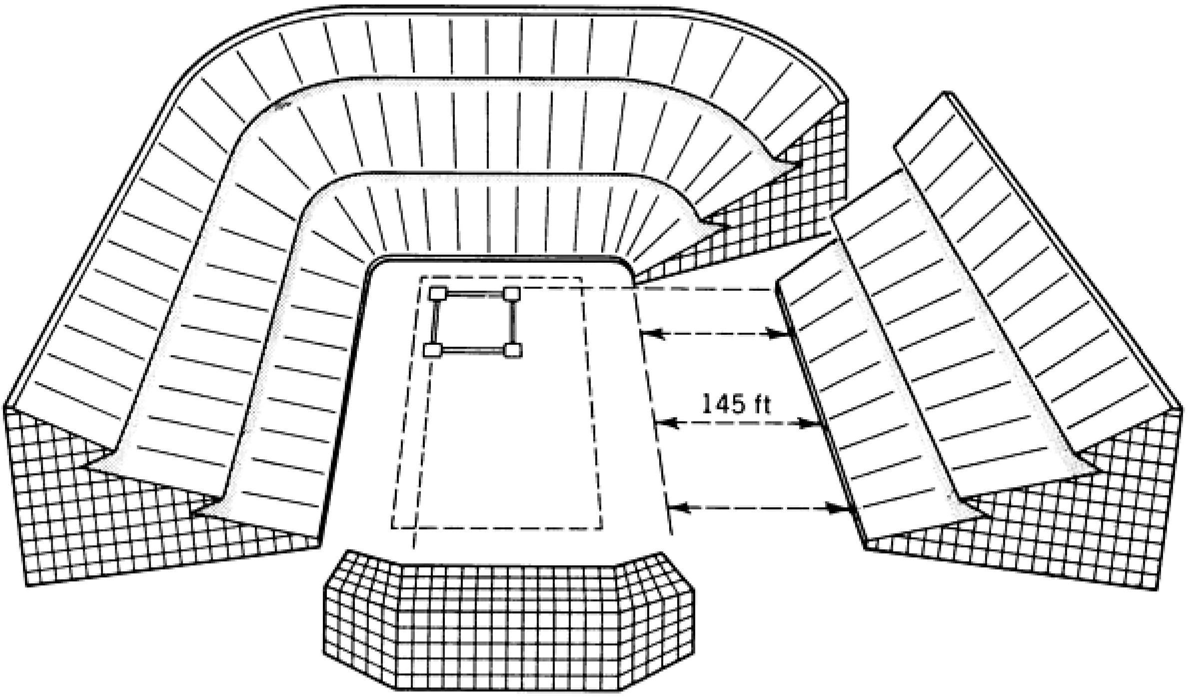


Fig. 35 Use of pressurized membrane bearings in Denver Mile High Stadium

It is surprising how little pressure is needed to support a large load. An available circular commercial unit 430 mm (17 in.) in diameter, with a pressure of about 21 kPa (3 psig) can carry a load of 270 kg (600 lb) with a flow of about 57 L/min (2 standard ft^3/min). The same unit at only 70 kPa (10 psig) can carry 900 kg (2000 lb) with, of course, a greater flow.

A novel application was made recently by a company in England that refurbishes full-size passenger railroad cars. A cradle platform was built to hold the four-wheel trucks at each end of the car. Then the platforms were floated on air-pressurized membrane bearings. Once this was done, the railroad car could be easily moved to any location on the factory floor without being encumbered by rails.

A few years ago, this type of hydrostatic compliant pad was offered as an option on a well-known brand of domestic refrigerator. By reversing the hose connection on a standard vacuum cleaner, the discharge pressure could be directed to the pads under the refrigerator and cause it to float in a film of air. It could then be moved very easily away from the wall, even when fully loaded, to allow for cleaning and maintenance.

Hydrostatic Journal Bearings

Externally pressurized journal bearings may be used when it is necessary to maintain a precise shaft position, with negligible friction, when rotational speed is insufficient to establish a hydrodynamic film. External pressurization is also used in combination with a hydrodynamic-type bearing to increase film stiffness, reduce the attitude angle, and raise the threshold of instability. The combination of these two actions produces what is known as a "hybrid bearing" (Fig. 36).

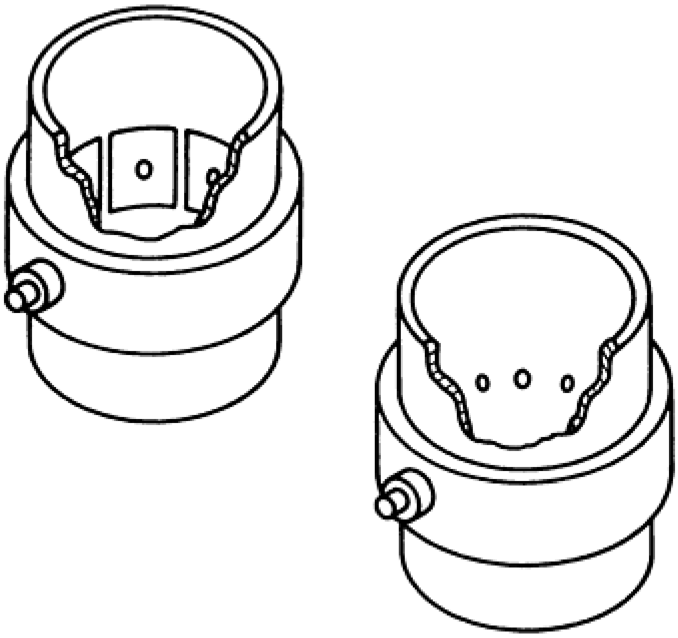


Fig. 36 Typical hybrid journal bearing designs. Source: Ref 9

The mutual contribution of each type of lubrication is of course of design decision. The rotor may be started or stopped without solid contact. Bearings for high-speed dental drills (500,000 rpm) are of the hybrid variety (Ref 46).

Externally pressurized journal bearings are affected by a great many combinations of parameters, so their design is relatively complex. Lund (Ref 47) has combined first-order self-acting and externally pressurized perturbation solutions for small eccentricity ratios. The results are found to be satisfactory and in good agreement with experimental results for small eccentricity ratios.

As with thrust bearings, the recesses must be kept as small as possible to avoid pneumatic instability. Multiple feed orifices or porous sections may be an answer to the problem. With multiple feeds, the assumption is usually made that the discrete points be considered as a continuous line source of pressure.

Comprehensive design charts for externally pressurized journal bearings are presented in Ref 48. Shapiro (Ref 49) has evaluated a three-sector hybrid journal bearing for both steady-state and dynamic behavior. He has published the results for this gas bearing, which uses orifice compensation, in tabular and graphical forms.

Many applications of externally pressurized gas bearings have been made to instruments, with remarkable increases in precision. Wunch, in Ref 45, describes the revolution that gas bearings have produced in the field of metrology. The surprising conclusion is that even though the components of the air bearing have been produced with normal manufacturing tolerances on dimensions and surface finishes, the rotational accuracy of the final instrument itself is of the order 0.075 to 0.125 μm. The explanation assumes the pressurized film has an averaging effect, and the small surface undulations of the bushing (surface roughness) are not transmitted to the shaft. This is similar to the smooth motion of a row boat on the surface of a pond; the boat is not influenced by the roughness of the bottom of the pond.

In machine tool applications, Lewis (Ref 50) has shown that grinder spindles, shapers, and gear cutters will actually show run-outs measured in microinches using externally pressurized journal bearings.

Materials for Gas-Lubricated Bearings

Engineers have been forced to search for materials that, when operated under extremely adverse conditions, will deliver acceptable life expectancy and reasonably low friction. These adverse conditions may be associated with high

temperature, with the use of lubricants with low viscosity and little or no natural oiliness (like gas), or with dry operation where no lubricants are present at all.

Experimentation has led in the direction of harder and harder materials, such as refractory materials and ceramics (Ref 51, 52, 53). These materials are usually brittle, so impact must be avoided. When they are used as rigid bearings with gas lubrication, the geometry must be precise and true because there is little possibility of conformability with such materials (as there could be with tin or lead-based babbitt, for example). Rigid designs, relatively light loads, and elimination of thermal or elastic deformations are necessary.

Investigations of gas-lubricated rigid thrust bearings and journal bearings operating under high temperature and high-speed conditions have shown favorable performance with the use of hard ceramic materials. Materials that might be considered for such applications include boron carbide, chromium carbide, silicon carbide, titanium carbide, alumina, silicon nitride, tungsten carbide, and chromium oxide.

Ceramics and cermets exhibit superior wear resistance. Cermets are ceramics that have been bonded with metals to improve their ability to handle impact and shock loading. The results, of course, vary with the ceramic and its bonding material. Cobalt is often used as a binder.

Hinkle and Fuller (Ref 54) conducted a study of the friction and wear of various materials for gas-lubricated rigid journal bearings under conditions of start-stop and high-speed whirl-induced rubbing. The tests were performed with a variety of imposed loads, temperatures, and ambient atmospheric conditions. Of the 36 combinations of materials investigated, by far the best performance was exhibited by aluminum oxide (alumina) against itself. The best specimens showed no visible evidence of wear even after extended testing.

Surprisingly, of the four different types of aluminum oxide used, one was far superior to the other three. The binder, the density, and the grain size are apparently strong variables. It might be advisable to get guarantees of friction and wear characteristics from a prospective vendor before selecting a particular grade of aluminum oxide.

Ceramics do provide low wear, but they also have relatively high friction. This may be a problem in some applications. In instrument gyroscopes, the electric motor drive is very small and light and has a low starting torque. Consequently, unless the friction of the bearing material is low enough, the motor will stall and not start. Much current research in gyro bearings is directed toward the reduction of friction in ceramic bearings through the introduction of some form of boundary lubrication.

Research carried out at the University of Rhode Island has determined that if final grinding of the ceramic is done in a bath of a prescribed boundary lubricant, a surface modification is achieved that provides lower friction. For silicon nitride, octadecanoic armide is recommended. For glass ceramics such as Pyroceram, the use of dioctedecyl disulfide is indicated.

In England, the British Royal Navy Scientific Service frequently uses silicon nitride against silicon nitride for gyro air bearings. They have been successful with small built-in graphite brushes that spread a thin film of solid graphite on the running surfaces. This technique not only reduces wear and lowers friction, but it also burnishes the surfaces to a high gloss while running. An alternative technique is burnishing the surfaces before installation with either molybdenum disulfide (MoS_2) or Teflon. Reducing friction with ceramics is an active research field.

Surface Coatings. High-density ultrahard coatings can be applied to the surfaces of ordinary commercial machine elements and provide them with a highly wear-resistant capability. Typical applications include parts for use in jet aircraft and in the chemical, textile, steel, and data processing industries. New coating technologies for tribological applications are described in Ref 55. Many different techniques can be used in making these surface depositions, depending on the end results desired.

Surface coating is a specialized field in which dramatic developments have been made within the last decade. For example, the metallic foil compliant-surface bearings, described earlier, have been coated with wear- and friction-resistant materials for use in various applications. Commercially bonded MoS_2, aluminum oxide, chromium oxide, and a three-component commercial blend of aluminum oxide, silicon oxide, and chromium oxide have all been used successfully for coating these bearings.

References

1. H.D. Smyth, *Atomic Energy for Military Purposes,* Princeton University Press, 1946
2. D.D. Fuller, Ed., *Proceedings, First International Symposium on Gas-Lubricated Bearings* (Washington, D.C.), ACR-49, 26-28 Oct 1959
3. *Proceedings, Second International Symposium on Gas Lubrication* (Las Vegas), 17-20 June 1968, American Society of Mechanical Engineers
4. W.A. Gross, Ed., *Fluid Film Lubrication ,* John Wiley & Sons, 1980
5. A.S. Raimondi, A Numerical Solution for the Gas-Lubricated, Full Journal Bearing of Finite Length, *Trans. ASME,* Apr 1961, p 131-155
6. W.A. Gross, A Gas Film Lubrication Study; Part I: Some Theoretical Analyses of Slider Bearings, *IBM J. Res. Dev.,* Vol 3, July 1959, p 237-255
7. G.W.K. Ford, D.M. Harris, and D. Pantall, Principles and Applications of Hydrodynamic-Type Gas Bearings, *Proc. Inst. Mech. Eng.,* Vol 171, 1957, p 93-113, discussion, p 113-128
8. O. Reynolds, On the Theory of Lubrication and Its Application to Mr. Beauchamp Towser's Experiments, Including an Experimental Determination of the Viscosity of Olive Oil, *Philos. Trans. R. Soc. (London) A,* Vol 177, Part I, 1886, p 157-234
9. D.D. Fuller, *Theory and Practice of Lubrication for Engineers,* 2nd ed., John Wiley & Sons, 1984
10. W.A. Gross, A Gas Film Lubrication Study; Part I: Some Theoretical Analyses of Slider Bearings, *IBM J. Res. Dev.,* Vol 3, July 1959, p 248
11. S. Abramovitz, Theory for a Slider Bearing with a Convex Pad Surface; Side Flow Neglected, *J. Franklin Inst.,* Vol 259 (No. 3), Mar 1955, p 221-233
12. R.K. Brunner, J.M. Harker, K.E. Haughton, and A.G. Osterlund, A Gas Film Lubrication Study; Part III: Experimental Investigation of Pivoted Slider Bearing, *IBM J. Res. Dev.,* No. 3, 1959, p 260-274
13. A. Kingsbury, Experiments with an Air Lubricated Journal, *J. Am. Soc. Naval Eng.,* Vol 9, 1897, p 267-292
14. H.F. Brubach, Some Laboratory Applications for the Low Friction Properties of the Dry Hypodermic Syringe, *Rev. Sci. Instrum.,* Vol 18, May 1947, p 363-366
15. H.G. Elrod and S.B. Malanoski, "Theory and Design Data for Continuous-Film, Self-Acting Journal Bearings of Finite Length," Report I-A 2049-13, Franklin Institute Laboratories for Research and Development, Nov 1960
16. H.G. Elrod and S.B. Malanoski, "Theory and Design Data for Continuous-Film, Self-Acting Journal Bearing of Finite Length," (Supplement to Report I-A 1049-13), Report I-A 2049-17, Franklin Institute Laboratories for Research and Development, June 1962
17. H.G. Elrod and A. Burgdorfer, Refinements of the Theory of the Infinitely Long, Self-Acting, Gas-Lubricated Bearings *Proceedings, First International Symposium on Gas-Lubricated Bearings,* ACR-49, U.S. Government Printing Office, Oct 1959, p 93-118
18. A.S. Raimondi, A Numerical Solution for the Gas-Lubricated, Full Journal Bearing of Finite Length, *Trans. ASME,* Apr 1961, p 131-155
19. A.C. Hagg, The Influence of Oil-Film Journal Bearings on the Stability of Rotating Machines, *J. Appl. Mech. (Trans. ASME),* Vol 68, 1946, p A211-A220. Discussion, Vol 69, Mar 1947, p A77-A78
20. V. Castelli and H.G. Elrod, Solution for the Stability Problem for 360 Degree, Self-Acting Gas-Lubricated Bearings, *J. Basic Eng. (Trans. ASME),* Vol 87, Mar 1965, p 199-212
21. E.J. Gunter, J.G. Hinkle, and D.D. Fuller, "Design Guide for Gas-Lubricated, Tilting-Pad Journal and Thrust Bearings with Special Reference to High-Speed Rotors," NYO-2512-1, U.S. Atomic Energy Commission, Eng. Development Branch, I-A 2393-3-1, Contract AT 30-1-2512, Nov 1964
22. E.J. Gunter, J.G. Hinkle, and D.D. Fuller, The Effects of Speed, Load, and Film Thickness on the Performance of Gas-Lubricated Tilting-Pad Journal Bearings, *Trans. ASLE,* Vol 7, 1964, p 353-365
23. J. Corniglion, G. Kilmister, H. Woodley and A.D. Richards, Theoretical Design and Practical Performance of Tilting-Pad Gas Bearings, Paper No. 19, *Gas Bearing Symposium Proceedings,* University of

Southampton, United Kingdom, Apr 1967

24. C. Mech, Some Practical Performance Aspects of the Design of Gas-Bearing Blowers and Some Performances of Industrial Machines, Paper No. 16, *Gas Bearing Symposium Proceedings,* University of Southampton, United Kingdom, Apr 1967
25. J.H. Dunn, "Inspection of Two Brayton Rotating Units after Extensive Endurance Testing," Report TM X-73569, Lewis Research Center, National Aeronautics and Space Administration, 1976
26. V. Castelli and J. Pirvics, Equilibrium Characteristics of Axial-Groove Gas-Lubricated Bearings, *J. Lubr. Technol. (Trans. ASME),* Vol 85, p 177-195
27. G.K. Fischer, J.L. Cherubim, and D.D. Fuller, "Some Instabilities and Operating Characteristics of High-Speed Gas-Lubricated Journal Bearings," Paper No. 58-A-231, American Society of Mechanical Engineers
28. V. Castelli and J.H. Vohr, Performance Characteristics of Herringbone Grooved Journal Bearings Operating at High Eccentricity Ratios and with Misalignment, Paper No. 14, *Proceedings of Third Gas Bearing Symposium,* University of Southampton, United Kingdom
29. S.B. Malanoski, Experiments on an Ultrastable Gas Journal Bearing, *J. Lubr. Technol. (Trans. ASME),* Vol 89, Oct 1967, p 433-438
30. S.R. Carfagno and J.T. McCabe, "Summary of Investigations of Entrance Effects in Circular Thrust Bearings," Report I-2049-24, Franklin Institute Research Laboratories, Defense Documentation Center Report, AD619966, 1965
31. J.T. McCabe, H.G. Elrod, S. Carfagno, and R. Colsher, Summary of Investigations of Entrance Effects of Circular Thrust Bearings, Paper 17, *Proceedings of Fourth Gas Bearing Symposium,* Vol 1, University of Southampton, United Kingdom
32. P.S. Moller, Radial Flow without Swirl between Parallel Disks Having Both Supersonic and Subsonic Regions, *J. Basic Eng. (Trans. ASME),* Vol 88, 1966, p 147-154
33. J.H. Vohr, A Study of Inherent Restriction Characteristics for Hydrostatic Gas Bearings, Paper 30, *Proceedings of Fourth Gas Bearing Symposium,* Vol 2, University of Southampton, United Kingdom
34. H. Mori, A Theoretical Investigation of Pressure Depression in Externally Pressurized Gas-Lubricated Circular Thrust Bearings, *J. Basic Eng. (Trans. ASME),* Vol 83, 1961, p 201-208
35. J.H. Laub, Evaluation of Externally-Pressurized Gas Pivot Bearings for Instruments, *Proceedings, First International Symposium on Gas-Lubricated Bearings,* ACR-49, U.S. Government Printing Office, Oct 1959, p 435-481
36. L. Licht, D.D. Fuller, and B.Sternlicht, Self-Excited Vibrations of an Air-Lubricated Thrust Bearing, *Trans. ASME,* Vol 80, 1958, p 411-414
37. H.J. Sneck, A Survey of Gas-Lubricated Porous Bearings, *J. Lubr. Technol, (Trans. ASME),* Oct 1968, p 804-809
38. E.P. Gargiulo and P.W. Gilmour, A Numerical Solution for the Design of Externally Pressurized Porous Gas Bearings: Thrust Bearings, *J. Lubr. Technol. (Trans. ASME),* 1968, p 810-817
39. H. Blok and J.J. van Rossum, Foil Bearing--New Departure in Hydrodynamic Lubrication, *Lubr. Eng.,* Vol 9, 1953, p 316-320
40. B.J. Patel and A. Cameron, The Foil Bearing, Paper 73, *Proceedings, Conference on Lubrication and Wear,* Institute of Mechanical Engineers (London), Oct 1957, p 219-223
41. W.A. Gross, Analysis and Design of Foil Bearings, Paper No. 23, *Proceedings, Gas Bearing Symposium,* University of Southampton, United Kingdom, Apr 1967
42. F.J. Suriano, R.D. Dayton, and F.G. Woessner, "Test Experience with Turbine-End Foil-Bearing-Equipped Gas Turbine Engines," Paper 83-GT-73, American Society of Mechanical Engineers, 1983
43. L. Licht, The Dynamic Characteristics of a Turborotor Simulator on Gas-Lubricated Foil Bearings, *J. Lubr. Technol. (Trans. ASME),* Vol 94, 1972, p 211-222
44. "Gas Lubricated Foil Bearing Development for Advanced Turbomachines," Report AF APL-TR-76-114, Vol I and II, Air Force Aero Propulsion Laboratory, 1977
45. S.B. Levy and C.H. Coogan, Flexible Membrane Hydrostatic Air Bearing *J. Lubr. Technol. (Trans. ASME),*

Vol 90, 1968, p 184-190
46. N.S. Grassam and J.W. Powell, *Gas Lubricated Bearings,* Butterworths, London, 1964
47. J.W. Lund, The Hydrostatic Gas Journal Bearing with Journal Rotation and Vibration, *J. Basic Eng. (Trans. ASME),* Vol 86, June 1964, p 328-336
48. D.F. Wilcock, Ed., *MTI Gas Bearing Design Manual,* Mechanical Technology Inc., 1972
49. W. Shapiro, Steady-State and Dynamic Analyses of Gas-Lubricated Hybrid Journal Bearings, *J. Lubr. Technol. (Trans. ASME),* Jan 1969, p 171-180
50. T.G. Lewis, "Hydrostatic Gas Bearings and Metrology in Ultra-Precision Machining," 1965 ASME Spring Lubrication Symposium, American Society of Mechanical Engineers, June 1965
51. E.L. Hemingway, Surface Finish--Key to Bearing Life, *Mach. Des.,* Jan 1945, p 168, 170
52. J.T. Burwell, Jr., Teaching Lubrication and Bearing Design, Part V, *Mach. Des. Manuf. Bull.,* Vol XV (No. 6), March 1949
53. R. Poppinga, *Wear and Lubrication of Piston Rings and Cylinders,* Society of Tribologists and Lubrication Engineers (ASLE), 1948
54. J.G. Hinkle and D.D. Fuller, Evaluation of Friction and Wear Characteristics of Materials for Gas-Lubricated Bearings under Conditions of Start-Stop and Whirl-Induced Rubbing Paper No. 24, *Proceedings,* University of Southampton, United Kingdom, Apr 1967, p 24-1 to 24-34
55. W. Winer and M. Peterson, Ed., *Wear Control Handbook,* ASME Research Committee on Lubrication, American Society of Mechanical Engineers, 1980

Friction, Lubrication, and Wear of Gears

Robert Errichello, GEARTECH

Introduction

BECAUSE GEARS are such common machine components, they may be taken for granted. It is not generally appreciated that they are complex systems requiring knowledge from all the engineering disciplines for their successful design. Gear design is a process of synthesis in which gear geometry, materials, heat treatment, manufacturing methods, and lubrication are selected to meet the requirements of a given application. The designer must design the gearset with adequate strength, wear resistance, and scuffing resistance. To do this, he or she must consider gear tribology. The choice of lubricant and its application method is as important as the choice of steel alloy and heat treatment. The interrelationship of the following factors must be considered:

- Gear tooth geometry
- Gear tooth motion (kinematics)
- Gear tooth forces (static and dynamic)
- Gear tooth material and surface characteristics (physical and chemical)
- Lubricant characteristics (physical and chemical)
- Environmental characteristics (physical and chemical)

Note: Originally published in *Lubrication Engineering,* Jan-April, 1990. Reprinted by permission of the Society of Tribologists and Lubrication Engineers. All rights reserved.

Gear Tooth Failure Modes

To obtain optimum minimum-weight gearsets, the gear designer must be aware of the intricate details of many competing modes of failure. In its nomenclature publication (Ref 1), the American Gear Manufacturers Association (AGMA) has classified 20 modes of gear failure under the broad categories of wear, surface fatigue, plastic flow, breakage, and associated gear failures. References 2, 3, 4, 5, 6 also give gear failure modes. For our purposes, the basic categories are overload, bending fatigue, Hertzian fatigue, wear, and scuffing (Table 1).

Table 1 Basic failure modes of gear teeth

Nonlubrication-related failures

Overload
Brittle fracture
Ductile fracture
Plastic deformation
    Cold flow
    Hot flow
    Indentation
        Rolling
        Bruising
        Peening
        Brinelling
    Rippling (fish scaling)
    Ridging
    Bending (yielding)
    Tip-to-root interference

Bending fatigue **Low-cycle fatigue ($\leq$1000 cycles to failure)** **High-cycle fatigue (>1000 cycles to failure)**

Lubrication-related failures

Hertzian fatigue
Pitting
    Initial
    Superficial
    Destructive
    Spalling
Micropitting
    Frosting
    Gray staining
    Peeling
Subcase fatigue (case crushing)

Wear
Adhesion
    Normal
    Running-in
    Mild
    Moderate
    Severe
    Excessive
Abrasion
    Scoring
    Scratching
    Plowing
    Cutting
    Gouging
Corrosion
Fretting corrosion
Cavitation
Electrical discharge damage
Polishing (burnishing)

Scuffing **Scoring**

Galling **Seizing** **Welding** **Smearing** **Initial** **Moderate** **Destructive**

Many gear failures are known by several names and qualifying terms, such as initial, moderate, destructive, and so on. The term "scoring" has been used in the past in the United States, while the term "scuffing" is used in Europe to describe the severe form of adhesive wear that involves the welding and tearing of the surfaces of gear teeth. To agree with current usage, the term scuffing will be used in this article when referring to this failure mode. The term scoring implies scratching, and it will be used to describe abrasive wear rather than scuffing.

Lubrication-Related Failure Modes

This article is concerned with gear tooth failures that are influenced by friction, lubrication, and wear. Pitting or scuffing may cause the gear teeth to deteriorate and generate dynamic forces, which in turn cause the gear teeth to fail by bending fatigue. In these cases, the bending failure is secondary and not directly related to lubrication, whereas pitting or scuffing are the primary failure modes, and both are definitely influenced by lubrication. The failure analyst must discern and difference between primary and secondary failure modes because the wrong corrective action is likely to be recommended if a secondary failure mode is mistaken for the primary failure mode. For example, increasing the size of the gear teeth to prevent reoccurrence of the above-mentioned bending failure would only make the situation worse by lowering the pitting and scuffing resistance. Godfrey (Ref 7) gives a good description of lubrication-related failure modes.

On the basis of the above considerations, overload and bending fatigue are judged to be unrelated to lubrication and are eliminated from further discussion together with another category, subcase Hertzian fatigue. Although corrosion, fretting-corrosion, cavitation, and electrical discharge damage are influenced by lubrication, they will not be discussed in this article because these failure modes occur relatively rarely in gear teeth. Hence, only the following failure modes are discussed in this article:

- Hertzian fatigue (including pitting and micropitting)
- Wear (including adhesion, abrasion, and polishing)
- Scuffing

Hertzian Fatigue. Pitting and micropitting are two lubrication-related gear tooth failure modes caused by Hertzian fatigue.

Pitting is a common failure mode for gear teeth because they are subjected to high Hertzian contact stresses and many stress cycles. For example, through-hardened gears are typically designed to withstand contact stresses of approximately 700 MPa (100 ksi), while the contact stresses can carburized gears may reach 2100 MPa (300 ksi). In addition, a given tooth on a pinion that is revolving at 3600 rev/min accumulates over 5×10^6 stress cycles every 24 h.

Pitting is a fatigue phenomenon (Ref 8) that occurs when a fatigue crack initiates either at the surface of the gear tooth or at a small depth below the surface. The crack usually propagates for a short distance in a direction roughly parallel to the tooth surface before turning or branching to the surface. When the cracks have grown to the extent that they separate a piece of the surface material, a pit is formed. If several pits grow together, the resulting larger pit is often referred to as a "spall." There is no endurance limit for Hertzian fatigue, and pitting occurs even at low stresses if the gears are operated long enough. Because there is no endurance limit, gear teeth must be designed for a suitable finite lifetime.

To extend the pitting life of a gearset, the designer must keep the contact stress low and the material strength and lubricant specific film thickness high. There are several geometric variables, such as diameter, face width, number of teeth, pressure angle, and so on, that can be optimized to lower the contact stress. Material alloys and heat treatment are selected to obtain hard tooth surfaces with high strength. Maximum pitting resistance is obtained with carburized gear teeth because they have hard surfaces, and carburizing induces beneficial compressive residual stresses that effectively lower the load stresses. The drawbacks to using carburized gear teeth are that they are relatively expensive to produce and that they must be finished by grinding. The details for obtaining high lubricant specific film thickness will be explained

later when elastohydrodynamic (EHD) lubrication is discussed (see the section "Elastohydrodynamic Lubrication" in this article), but general recommendations are to use an adequate supply of cool, clean, and dry lubricant that has adequate viscosity and a high pressure-viscosity coefficient.

Pitting may initiate at the surface or at a subsurface defect, such as a nonmetallic inclusion. With gear teeth, pits are most often of the surface-initiated type because the lubricant film thickness is usually low, resulting in relatively high metal-to-metal contact. The interaction between asperities or contacts at defects, such as nicks or furrows, creates surface-initiated, rather than subsurface-initiated cracks. For high-speed gears with smooth surface finishes, the film thickness is greater and subsurface-initiated pitting, rather than surface-initiated, may predominate. In these cases, pitting usually starts at a subsurface inclusion, which acts as a point of stress concentration. Cleaner steels, such as those produced by vacuum melting, prolong the pitting life by reducing the number of inclusions.

Contamination from water in the lubricant is believed to promote pitting through hydrogen embrittlement of the metal, and abrasive particles in the lubricant cause pitting by indenting the tooth surfaces, causing stress concentrations, and disrupting the lubricant film. At present, the influence of lubricant additives on pitting is unresolved.

The following recommendations serve as guidelines for preventing the onset of pitting in gearsets:

- Reduce contact stresses by reducing loads or optimizing gear geometry
- Use clean steel, properly heat treated to high hardness, preferably by carburizing
- Use smooth tooth surfaces produced by careful grinding or honing
- Use an adequate amount of cool, clean, and dry lubricant of adequate viscosity

Micropitting. On relatively soft gear tooth surfaces, such as those of through-hardened gears, Hertzian fatigue forms large pits with dimensions on the order of millimeters. With surface-hardened gears (for example, carburized, nitrided, induction hardened, and flame hardened), pitting may occur on a much smaller scale, typically only 10 μm (400 μin.) deep. To the naked eye, the areas where micropitting has occurred appear frosted, and "frosting" is a popular term for micropitting. Japanese researchers (Ref 9) have referred to the failure mode as "gray staining" because the light-scattering properties of micropitting give the gear teeth a gray appearance. With scanning electron microscopy (SEM), it is immediately evident that micropitting proceeds by the same fatigue process as classical pitting, except the pits are extremely small.

In many cases, micropitting is not destructive to the gear tooth surface. It sometimes occurs only in patches and may stop after the tribological conditions have been improved by running-in. The micropits may actually be removed by mild polishing wear during running-in, in which case the micropitting is said to "heal." However, there have been examples (Ref 9, 10, 11) where micropitting has escalated into full-scale pitting, leading to the destruction of the gear teeth.

The specific film thickness is the most important parameter that influences micropitting. Damage seems to occur most readily on gear teeth with rough surfaces, especially when they are lubricated with low-viscosity lubricants. Gears finished with special grinding wheels to a mirrorlike finish (Ref 12) have effectively eliminated micropitting. Slow-speed gears are prone to micropitting because their film thickness is low.

The prevent micropitting, the specific film thickness should be maximized by using smooth gear tooth surfaces, high-viscosity lubricants, and high speeds. Experiments (Ref 10) have shown that flame-hardened and induction-hardened gears have less resistance to micropitting than carburized gears of the same hardness. This is probably due to the lower carbon content of the surface layers of the flame-hardened and induction-hardened gears.

The following recommendations serve as guidelines for preventing the onset of micropitting in gearsets:

- Use smooth tooth surfaces produced by careful grinding or honing
- Use an adequate amount of cool, clean, and dry lubricant of the highest viscosity permissible
- Use high speeds if possible
- Use carburized steel with proper carbon content in the surface layers

Wear. Gearsets are susceptible to wear caused by adhesion, abrasion, and polishing.

Adhesion. Adhesive wear is classified as "mild" if it is confined to the oxide layers of the gear tooth surfaces. If, however, the oxide layers are disrupted and bare metal is exposed, the transition to severe adhesive wear usually occurs. Severe adhesive wear is termed scuffing and will be discussed in the section "Scuffing" in this article. Here we assume that scuffing has been avoided through proper design of the gears, selection of the lubricant, and control of the running-in process.

When new gear units are first operated, the contact between the gear teeth is not optimum because of unavoidable manufacturing inaccuracies. If the tribological conditions are favorable, mild adhesive wear occurs during running-in and usually subsides with time, resulting in a satisfactory lifetime for the gears. The wear that occurs during running-in is beneficial if it smooths the tooth surfaces (thereby increasing the specific film thickness) and if it increases the area of contact by removing minor imperfections through local wear. To ensure that the wear rate remains under control, new gearsets should be run-in by being operated for at least the first ten hours at one-half load.

The amount of wear considered tolerable depends on the expected lifetime for the gears and requirements for control of noise and vibration. Wear is considered excessive when the tooth profiles wear to the extent that high dynamic loads occur or the tooth thickness is reduced to the extent that bending fatigue becomes possible.

Many gears, because of practical limits on lubricant viscosity, speed, and temperature, must operate under boundary-lubricated conditions in which some wear is inevitable. Highly loaded, slow speed (<30 m/min, or 100 ft/min), boundary-lubricated gears are especially prone to excessive wear. Tests with slow-speed gears (Ref 10) have shown that nitrided gears have good wear resistance, whereas carburized and through-hardened gears have similar but lower wear resistance. Winter and Weiss (Ref 10) concluded that lubricant viscosity has the greatest effect on slow-speed adhesive wear, and that high-viscosity lubricants reduce the wear rate significantly. Winter and Weiss (Ref 10) also found that sulfur-phosphorus additives can be detrimental with slow-speed (<3 m/min, or 10 ft/min) gears, giving very high wear rates.

A few gear units operate under ideal conditions with smooth tooth surfaces, high pitch line speed, and high lubricant film thickness. For example, turbine gears that operated almost continuously at 9000 m/min (30,000 ft/min) pitch line speed still had the original machining marks on their teeth, even after operating for 20 years. Most gears, however, operate between the boundary and full-film lubrication regimes, under EHD conditions. In the EHD regime, with the proper type and viscosity of lubricant, the wear rate usually reduces during running-in and adhesive wear virtually ceases once running-in is completed. If the lubricant is properly maintained (cool, clean, and dry), the gearset should not suffer an adhesive wear failure.

The following recommendations serve as guidelines for preventing the onset of adhesive wear in gearsets:

- Use smooth tooth surfaces
- If possible, run-in new gearsets by operating the first 10 hours at one-half load
- Use high speeds if possible. Otherwise, recognize that highly loaded slow-speed gears are boundary lubricated and are especially prone to excessive wear. For these conditions, specify nitrided gears and the highest permissible lubricant viscosity
- For very slow-speed gears (<3 m/min, or 10 ft/min), avoid using lubricants with sulfur-phosphorus additives
- Use an adequate amount of cool, clean, and dry lubricant of the highest viscosity permissible

Abrasion. Abrasive wear on gear teeth is usually caused by contamination of the lubricant by hard, sharp-edged particles. Contamination enters gearboxes by being built-in, internally generated, ingested through breathers and seals, or inadvertently added during maintenance.

Many gear manufacturers do not fully appreciate the significance of clean assembly; it is not uncommon to find sand, machining chips, grinding dust, weld splatter, or other debris in new gearboxes. To remove built-in contamination, the gearbox lubricant should be drained and flushed before start-up and again after the first 50 h of operation, refilled with the recommended lubricant, and a new oil filter should be installed.

Internally generated particles are usually wear debris from gears or bearings due to Hertzian fatigue pitting or adhesive and abrasive wear. The wear particles are especially abrasive because they become work-hardened when they are trapped between the gear teeth. Internally generated wear debris can be minimized by using accurate surface-hardened gear teeth (with high pitting resistance), smooth surfaces, and high-viscosity lubricants.

Breather vents are used on gearboxes to vent internal pressure, which may occur when air enters through seals, or when air within the gearbox expands (or contracts) during the normal heating and cooling of the gear unit. The breather vent should be located in a clean, nonpressurized area and should have a filter to prevent ingress of airborne contaminants. In especially harsh environments, the gearbox can be completely sealed, and the pressure variation can be accommodated by an expansion chamber with a flexible diaphragm.

All maintenance procedures that involve opening any part of the gearbox or lubrication system must be carefully performed to prevent contamination of the gearbox system.

Abrasive wear due to foreign contaminants, such as sand or internally generated wear debris, is called three-body abrasion and is a common occurrence. Two-body abrasion also occurs when hard particles or asperities on one gear tooth abrade the opposing tooth surface. Unless the tooth surfaces of a surface-hardened gear are smoothly finished, they will act like files if the mating gear is appreciably softer. This is the reason that a worm pinion is polished after grinding before it is run with a bronze worm wheel. Manufacturers of computer disk drives have found that stainless steel pinions mated with anodized aluminum racks have excessively high wear rates. The anodized layer of the aluminum rack is extremely thin and brittle, and it breaks up and impregnates the relatively soft stainless steel pinion. The aluminum oxide particles then act like emery paper and wear the teeth of the rack very quickly.

The lubrication system should be carefully maintained and monitored to ensure that the gears receive an adequate amount of cool, clean, and dry lubricant. For circulating-oil systems, fine filtration removes contamination. Filters as fine as 3 μm (120 μin.) have significantly increased gear life. For oil bath gearboxes, the lubricant should be changed frequently to remove contamination. Under normal operating conditions, the lubricant should be changed at least every 2500 h of operation or every six months, whichever occurs first. For critical gearboxes, a regular program of lubricant monitoring can help prevent gear failures by showing when maintenance is required. The lubricant monitoring should include spectrographic and ferrographic analysis of contamination, along with analysis of acid number, viscosity, and water content.

In summary, the following guidelines should be observed to prevent abrasive wear in gearsets:

- Remove built-in contamination from new gearboxes by draining and flushing the lubricant before start-up and again after the first 50 h of operation. Refill with the recommended lubricant and install a new filter
- Minimize internally generated wear debris by using surface-hardened gear teeth, smooth tooth surfaces, and high-viscosity lubricants
- Minimize ingested contamination by maintaining oil-tight seals and using filtered breather vents located in clean, nonpressurized areas
- Minimize contamination that is added during maintenance by using good housekeeping procedures
- For circulating-oil systems, use fine filtration
- For oil bath systems, change the lubricant at least every 2500 h or every six months
- Monitor the lubricant with spectrographic and ferrographic analysis together with analysis of acid number, viscosity, and water content

Polishing Wear. If the extreme-pressure (EP) antiscuff additives in the lubricant are too chemically reactive, they may cause polishing of the gear tooth surfaces until they attain a bright mirror finish. Although the polished gear teeth may look good, polishing wear is undesirable because it generally reduces gear accuracy by wearing the tooth profiles away from their ideal form. Antiscuff additives used in lubricants to prevent scuffing, such as sulfur and phosphorus, will be covered when scuffing is discussed. They function by forming iron-sulfide and iron-phosphate films on areas of the gear teeth where high temperatures occur. Ideally, the additives should react only at temperatures where there is a danger of welding. If the rate of reaction is too high, and there is a continuous removal of the surface films caused by a very fine abrasives in the lubricant, the polishing wear may be excessive (Ref 13).

Polishing wear can be prevented by using less chemically active additives. As an alternative to sulfur-phosphorus additives, antiscuff lubricants are available with dispersions of potassium borate (Ref 14) that deposit EP films without chemically reacting with the metal. Removing the abrasives in the lubricant by using fine filtration or frequent oil changes is helpful.

In summary, the following guidelines should be observed to prevent polishing wear in gearsets:

- Use less chemically active antiscuff additives (for example, borate)
- Remove abrasives from the lubricant by using fine filtration or frequent oil changes

Scuffing is defined as localized damage caused by solid-phase welding between sliding surfaces. It is accompanied by the transfer of metal from one surface to another due to welding and tearing. It may occur in any sliding and rolling contact where the oil film is not thick enough to separate the surfaces. The symptoms of scuffing are microscopically rough, matte, and torn surfaces. Surface analysis that shows transfer of metal from one surface to the other is proof of scuffing.

Scuffing can occur in gear teeth when they operate in the boundary lubrication regime. If the lubricant film is insufficient to prevent significant metal-to-metal contact, the oxide layers that normally protect the gear tooth surfaces may be broken through, and the bare metal surfaces may weld together. The sliding that occurs between gear teeth results in tearing of the welded junctions, metal transfer, and catastrophic damage.

In contrast to pitting and bending fatigue, which only occur after a period of running time, scuffing may occur immediately upon start-up. In fact, gears are most vulnerable to scuffing when they are new and their tooth surfaces have not yet been smoothed by running-in. For this reason, it is wise to run-in a new gearbox under one-half load for at least 10 h to reduce the surface roughness of the teeth before applying full load. The gear teeth can be coated with iron-manganese phosphate or plated with copper or silver to protect them from scuffing during the critical running-in period.

The basic mechanism of scuffing is not clearly understood, but by general agreement it is believed to be caused by intense frictional heating generated by the combination of high sliding velocity and intense surface pressure. Blok's (Ref 15) critical temperature theory is believed to be the best criterion for predicting scuffing. It states that scuffing will occur in gear teeth that are sliding under boundary-lubricated conditions when the maximum contact temperature of the gear teeth reaches a critical magnitude. For mineral oils without antiscuff/EP additives, each combination of oil and rubbing materials has a critical scuffing temperature that is constant, regardless of the operating conditions (Ref 16). The critical scuffing temperatures are not constant for synthetic lubricants and lubricants with antiscuff additives; they must be determined from tests that closely simulate the operating conditions of the gears.

Today, most antiscuff additives are sulfur-phosphorus compounds, which form boundary-lubricating films by chemically reacting with the metal surfaces of the gear teeth at local points of high temperature. Antiscuff films help prevent scuffing by forming solid films on the gear tooth surfaces and inhibiting true metal-to-metal contact. The films of iron sulfide and iron phosphate have high melting points, allowing them to remain as solids on the gear tooth surfaces even at high contact temperatures. The rate of reaction of the antiscuff additives is greatest where the gear tooth contact temperatures are highest. Because of the rubbing action of the gear teeth, the surface films are repeatedly scraped off and reformed. In effect, scuffing is prevented by substituting mild corrosion in its place. Occasionally, antiscuff additives (for example, sulfur) are too chemically active, causing polishing wear and necessitating a change to less aggressive additives. Lubricants with antiscuff additives of potassium borate do not cause polishing wear because they deposit glasslike boundary films without reacting with the metal.

For mineral oils without antiscuff additives, the critical scuffing temperature increases with increasing viscosity and ranges from 150 to 300 °C (300 to 570 °F). The increased scuffing resistance of high-viscosity lubricants is believed to be due to differences in chemical composition rather than increases in viscosity. However, a viscosity increase also helps to reduce the risk of scuffing by increasing the lubricant film thickness and reducing the contact temperature generated by metal-to-metal contact.

Scuffing is controlled by the total contact temperature, T_c, which consists of the sum of the gear bulk temperature, T_b, and the flash temperature, T_f:

$$T_c = T_b + T_f \qquad \textbf{(Eq 1)}$$

The bulk temperature is the equilibrium temperature of the surface of the gear teeth before they enter the meshing zone. The flash temperature is the local and instantaneous temperature rise that occurs on the gear teeth due to the frictional heating as they pass through the meshing zone.

Anything that reduces either the bulk temperature or the flash temperature will reduce the total contact temperature and lessen the risk of scuffing. Higher viscosity lubricants or smoother tooth surfaces help by increasing the specific film thickness, which in turn reduces the frictional heat and, therefore, the flash temperature. Also, the lubricant performs the important function of removing heat from the gear teeth. The lubricant must be supplied to the gear teeth in such a way that it removes heat rapidly and maintains a low bulk temperature. A heat exchanger can be used with a circulating-oil system to cool the lubricant before it is sprayed at the gears.

The gear designer can maximize scuffing resistance by optimizing the gear geometry so that the gear teeth are as small as possible, consistent with bending strength requirements, to reduce the temperature rise caused by sliding. Figure 1 shows that the rolling velocity of the pinion, V_{r1}, and the rolling velocity of the gear, V_{r2}, linearly increase from zero at the interference points to a maximum at each end of the path of contact. The sliding velocity is represented by the distance between the V_{r1} and V_{r2} lines. The amount of sliding is proportional to the distance from the pitch point, *P*, and is zero when the gear teeth contact at the pitch point, and largest at the ends of the path. Addendum modification can be used to balance and minimize the temperature rise that occurs in the addendum and dedendum of the gear teeth. The temperature rise may also be reduced by modifying the tooth profiles with slight tip and/or root relief to ease the load at the start and end of the engagement path where the sliding velocities are the greatest. Also, the gear teeth must be accurate and held rigidly in good alignment to minimize tooth loading and, therefore, the temperature rise.

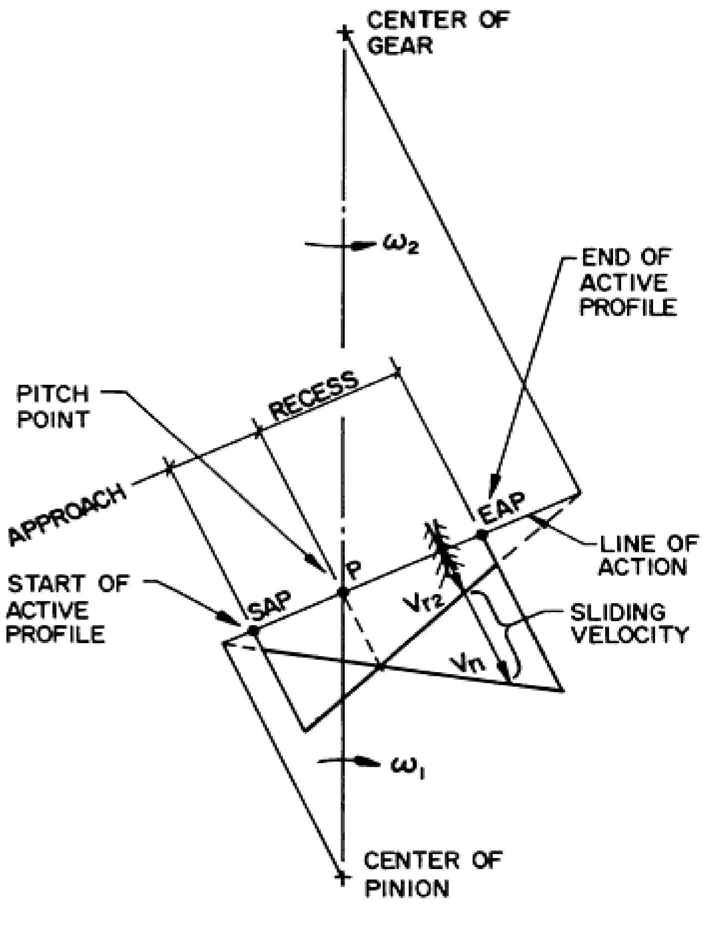


Fig. 1 Graphical representation of rolling velocities of pinion (V_{r1}) and gear (V_{r2}) in relation to sliding velocity

Gear materials should be chosen with their scuffing resistance in mind. Nitrided steels, such as Nitralloy 135M, are generally found to have the highest resistance to scuffing, whereas stainless steels are liable to scuff even under near-zero loads. The thin oxide layer on stainless steel is hard and brittle and breaks up easily under sliding loads, exposing the bare metal and thus promoting scuffing. Like stainless steel, anodized aluminum has a low scuffing resistance. Hardness does not seem to be a reliable indication of scuffing resistance.

In summary, the following guidelines should be observed to prevent scuffing in gearsets:

- Use smooth tooth surfaces produced by careful grinding or honing
- Protect the gear teeth during the critical running-in period by coating them with iron-manganese phosphate or plating them with copper or silver. Run-in new gearsets by operating the first 10 h at one-half load
- Use high-viscosity lubricants with antiscuff additives, such as sulfur, phosphorus, or borate
- Cool the gear teeth by supplying an adequate amount of cool lubricant. For circulating-oil systems, use a heat exchanger to cool the lubricant
- Optimize the gear tooth geometry by using small teeth, addendum modification, and profile modification
- Use accurate gear teeth, rigid gear mountings, and good helix alignment
- Use nitrided steels for maximum scuffing resistance. Do not use stainless steel or aluminum for gears if there is a risk of scuffing

Elastohydrodynamic Lubrication

Gear teeth are subjected to enormous contact pressures on the order of the ultimate tensile strength of hardened steel, yet they are quite successfully lubricated with oil films that are <1 μm (<40 μin.) thick. This is possible because lubricants have a fortuitous property that causes their viscosity to increase dramatically with increased pressure. Figure 2 depicts the region of contact between mating gear teeth. It shows the shape of the elastically deformed teeth and the pressure distribution developed within the contact zone. The molecular adsorption of the lubricant onto the gear tooth surfaces causes it to be dragged into the inlet region of the contact, where its pressure is increased due to the convergence of the tooth surfaces. The viscosity increase of the lubricant caused by the increasing pressure helps to entrain the lubricant into the contact zone. Once it is within the high-pressure Hertzian region of the contact, the lubricant cannot escape because its viscosity has increased to the extent where the lubricant is virtually a rigid solid.

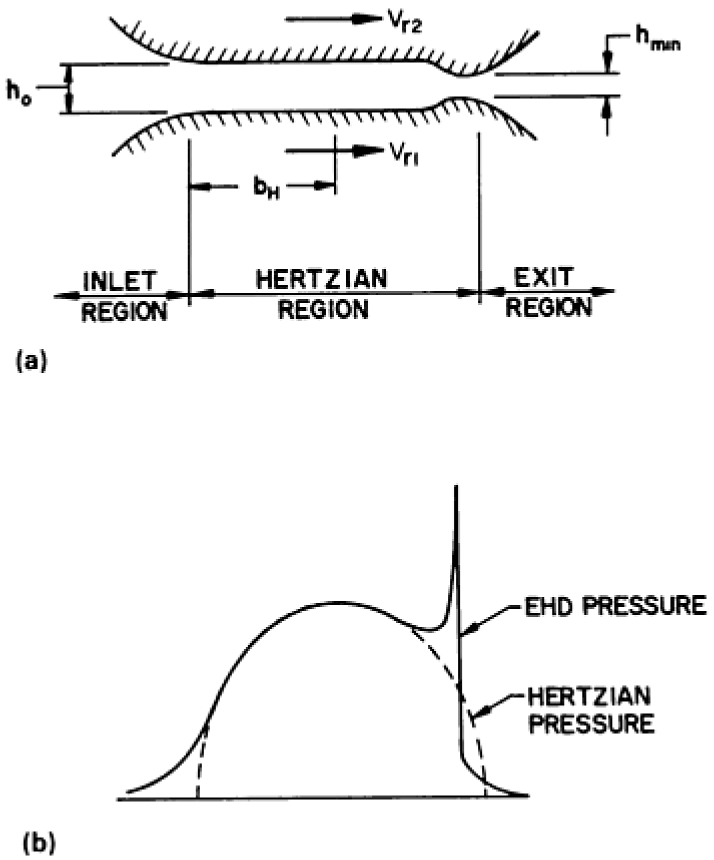


Fig. 2 Regions of elastohydrodynamic contact between the mating gear teeth of a gearset. (a) Schematic showing three distinct regions on pinion and gear tooth surfaces and key parameters determining oil film lubrication. (b) Plot of pressure distribution within contact zone. b_H, semiwidth of Hertzian contact band; h_0, central film thickness; h_{min}, minimum film thickness

The following equation, from Dowson and Higginson (Ref 17) gives the minimum film thickness that occurs near the exit of the contact. The minimum film thickness, h_{min}, is obtained from:

$$h_{min} = \frac{1.63\alpha^{0.54}\,(\mu_0 V_e)^{0.7}\,\rho_n^{0.43}}{(X_\Gamma w_{Nr})^{0.13}\,E_r^{0.03}} \qquad \textbf{(Eq 2)}$$

where α is the pressure-viscosity coefficient (in.^{2}/lbf), μ_0 is the absolute viscosity (reyns, or lbf · s/in.^{2}), and V_e is the entraining velocity, in which:

$$V_e = V_{r1} + V_{r2} \qquad \textbf{(Eq 3)}$$

where V_{r1} and V_{r2} are the rolling velocities of pinion and gear, respectively, given by:

$$V_{r1} = \omega_1\,\rho_2 \qquad \textbf{(Eq 4a)}$$

$$V_{r2} = \omega_1\,\rho_2 \qquad \textbf{(Eq 4b)}$$

in which ω_1 and ω_2 are the angular velocities of pinion and gear, respectively, and ρ_1 and ρ_2 are the transverse radii of curvature of the pinion and gear, respectively.

The normal relative radius of curvature, ρ_n, is given by:

$$\rho_n = \frac{\rho_1 \rho_2}{(\rho_2 \pm \rho_1)\cos \psi_b} \qquad \textbf{(Eq 5)}$$

where ψ_b is the base helix angle, $X\Gamma$ is the load-sharing factor, and ω_{Nr} is the normal unit load given by:

$$w_{Nr} = \frac{W_{Nr}}{L_{min}} \qquad \textbf{(Eq 6)}$$

in which W_{Nr} is the normal operating load and L_{min} is the minimum contact length.

E_r is the reduced modulus of elasticity, given by:

$$E_r = 2\left(\frac{1-\nu_1^2}{E_1} + \frac{1-\nu_2^2}{E_2}\right)^{-1} \qquad \textbf{(Eq 7)}$$

where v_1 is the Poisson's ratio of pinion, v_2 is the Poisson's ratio of gear, E_1 is the modulus of elasticity of pinion, and E_2 is the modulus of elasticity of gear.

Figure 3 gives average values of absolute viscosity, μ_0, versus bulk temperature for typical mineral gear lubricants with a viscosity index of 95.

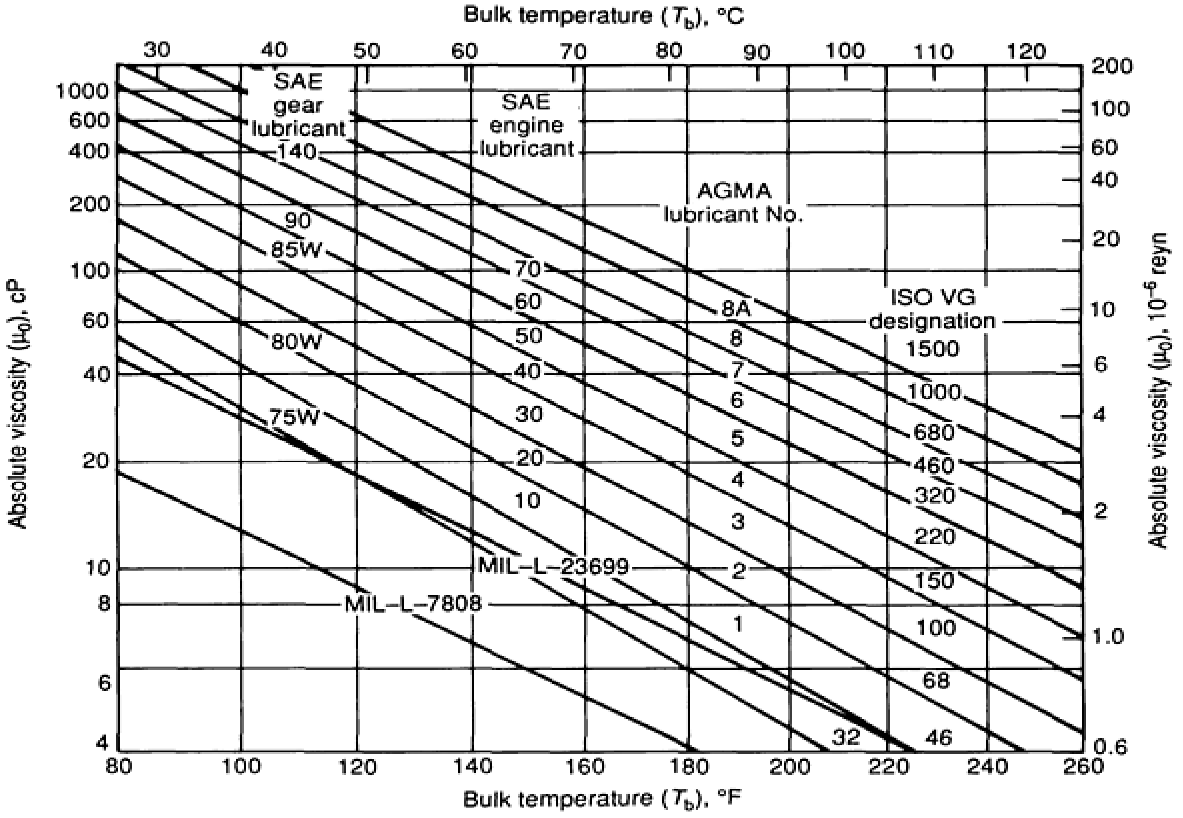


Fig. 3 Plot of absolute viscosity versus bulk temperature for selected mineral oil gear lubricants having a viscosity index of 95

The pressure-viscosity coefficient, α, ranges from 0.5×10^{-4} to 2×10^{-4} in.^{2}/lbf for typical gear lubricants. Data for pressure-viscosity coefficients versus bulk temperature for typical gear lubricants are given in Fig. 4.

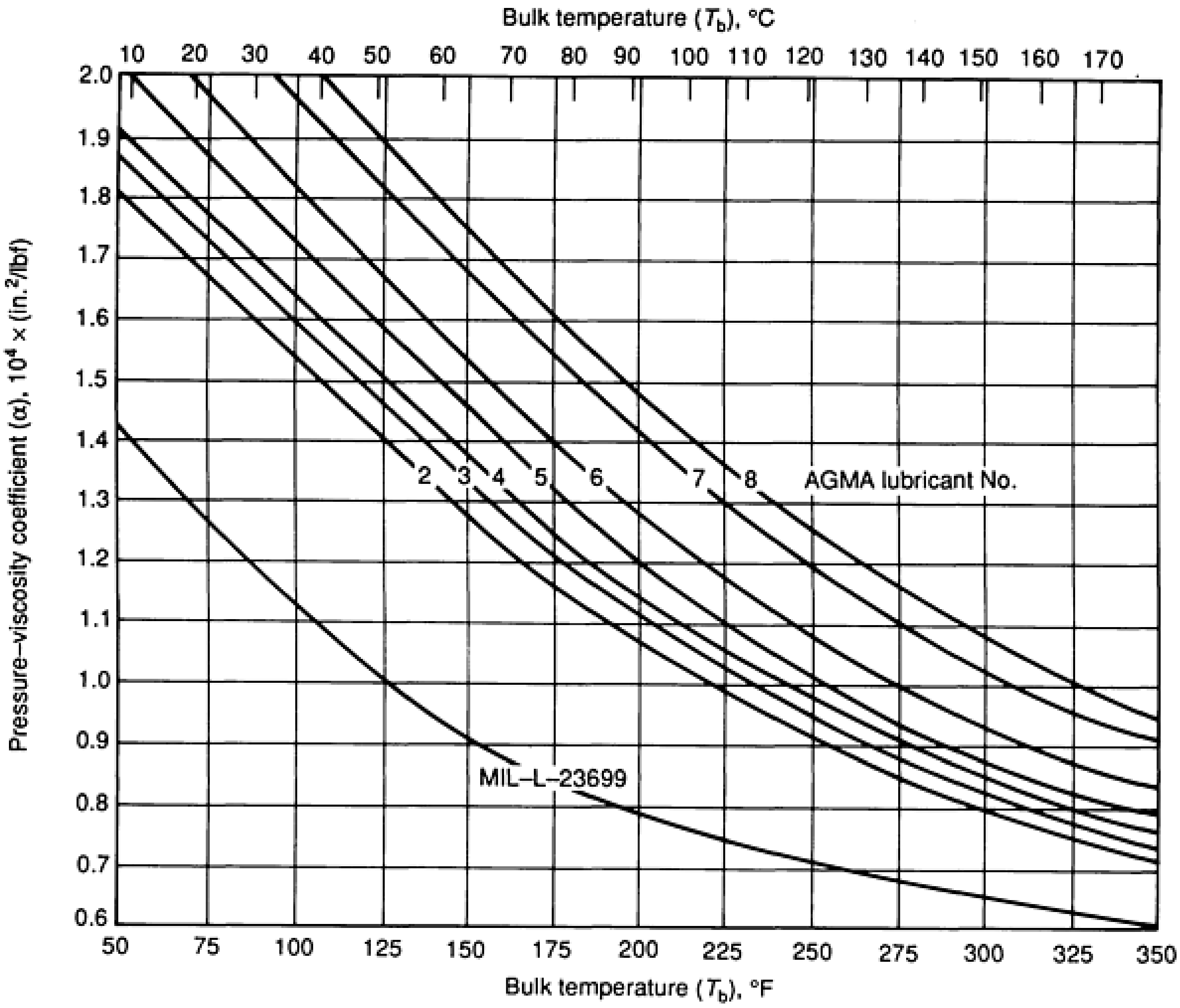


Fig. 4 Plot of pressure-viscosity coefficient versus bulk temperature for selected mineral oil gear lubricants

The specific film thickness, λ, is given by:

$$\lambda = \frac{h_{min}}{\sigma} \quad \textbf{(Eq 8)}$$

where h_{min} is the minimum film thickness and σ is the composite surface roughness given by:

$$\sigma = (\sigma_1^2 + \sigma_2^2)^{1/2} \quad \textbf{(Eq 9)}$$

in which σ_1 is the root-mean-square (rms) surface roughness of the pinion and σ_2 is the rms surface roughness of the gear.

Load-Sharing Factor. The load-sharing factor, $X\Gamma$, accounts for load sharing between succeeding pairs of teeth as affected by profile modification (tip and root relief) and whether the pinion or gear is the driver. Figure 5 gives plots of the load-sharing factors for unmodified and modified tooth profiles.

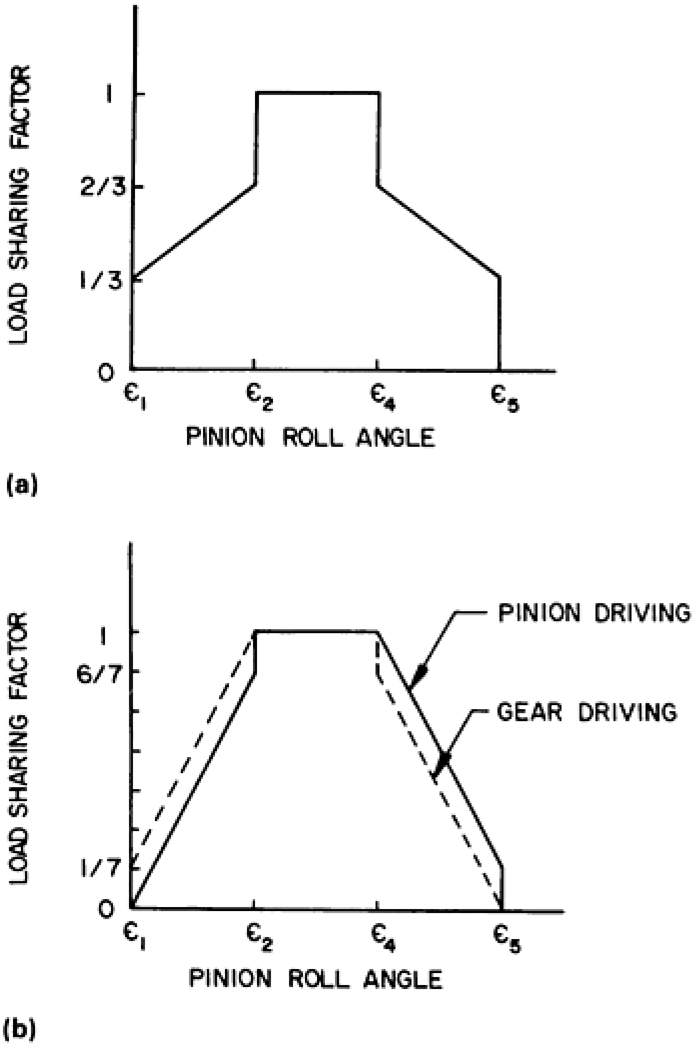


Fig. 5 Plot of load-sharing factor ($X\Gamma$) versus pinion roll angle. (a) Unmodified tooth profiles. (b) Modified tooth profiles.

As shown by the exponents in Eq 2, the film thickness is essentially determined by the entraining velocity, lubricant viscosity, and pressure-viscosity coefficient; the elastic properties of the gear teeth and the load have relatively small influences. In effect, the relatively high stiffness of the oil film makes it insensitive to load, and an increase in load simply increases the elastic deformation of the tooth surfaces and widens the contact area, rather than decreasing the film thickness.

Blok's contact temperature theory (Ref 15) states that scuffing will occur in gear teeth which are sliding under boundary-lubricated conditions when the maximum contact temperature of the gear teeth reaches a critical magnitude. The contact temperature is the sum of two components: the bulk temperature and the flash temperature, as described in Eq 1.

Blok's flash temperature equation as formulated in AGMA 2001-B88, Appendix A (Ref 18) for spur and helical gears is:

$$T_f = \frac{0.8\mu_m X_\Gamma w_{Nr} \left| (V_{r1})^{0.5} - (V_{r2})^{0.5} \right|}{B_M (b_H)^{0.5}} \quad \text{(Eq 10)}$$

where μ_m is the mean coefficient of friction, B_M is the thermal contact coefficient, and b_H is the semiwidth of the Hertzian contact band.

Mean Coefficient of Friction. The following equation gives a typical value of $0.06 < \mu_m < 0.18$ for the mean coefficient of friction for gears operating in the partial EHD regime ($\lambda < 1$). It may give values too low for boundary-lubricated gears, where μ_m may be greater than 0.2, or too high for gears in the full-film regime ($\lambda > 2$), where μ_m may be less than 0.01:

$$\mu_m = 0.06\left(\frac{50}{50 - S}\right) \quad \text{(Eq 11a)}$$

where:

$$\left(\frac{50}{50 - S}\right) \leq 3.0 \quad \text{(Eq 11b)}$$

in which S is the average rms surface roughness given by:

$$S = \frac{\sigma_1 + \sigma_2}{2} \quad \text{(Eq 12)}$$

Thermal Contact Coefficient. The thermal contact coefficient, B_M, is given by:

$$B_M = (\lambda_M \rho_M c_M)^{0.5} \quad \text{(Eq 13)}$$

where λ_M is the heat conductivity, ρ_M is the density, and c_M is the specific heat per unit mass. For typical gear steels, B_M is ~43 lbf/(in. · $s^{0.5}$ · °F).

Semiwidth of the Hertzian Contact Band. The semiwidth of the Hertzian contact band, b_H, is defined as:

$$b_H = \left(\frac{8X_\Gamma w_{Nr} \rho_n}{\pi E_r}\right)^{0.5} \quad \text{(Eq 14)}$$

Bulk Temperature. The gear bulk temperature, T_b, is the equilibrium bulk temperature of the gear teeth before they enter the meshing zone. In some cases, the bulk temperature may be significantly higher than the temperature of the oil supplied to the gear mesh. In a test with ultrahigh-speed gears (Ref 19), the pinion bulk temperature was 135 °C (275 °F), which is 95 °C (171 °F) hotter than the oil inlet temperature. For turbine gears at lower speeds, the bulk temperature rise of the gear teeth over the inlet oil temperature may range from 11 °C (20 °F) at 3700 m/min (1.2×10^4 ft/min) pitch line velocity to 22 °C (40 °F) at 4900 m/min (1.6×10^4 ft/min). At similar speeds, the bulk temperature rise of aircraft gears with less oil flow may range from 22 to 33 °C (40 to 60 °F).

Scuffing Temperature. The scuffing temperature, T_s, is the contact temperature at which scuffing is likely to occur with the chosen combination of lubricant and gear materials.

For mineral oils without antiscuff additives or for mineral oils with low concentrations of antiscuff additives, the scuffing temperature is independent of the operating conditions for a fairly wide range. For these oils, the scuffing temperature may be correlated with the composition of the oil. The viscosity grade is a convenient index of the composition and, thus, of the scuffing temperature.

For nonantiscuff mineral oils, the mean scuffing temperature (50% chance of scuffing) in °F is given by:

$$T_s = 146 + 59 \ln v_{40} \qquad \textbf{(Eq 15)}$$

where v_{40} is the kinematic viscosity at 40 °C (105 °F) in centistokes.

For mineral oils with low concentrations of antiscuff additives, the mean scuffing temperature in °F is given by:

$$T_s = 245 + 59 \ln v_{40} \qquad \textbf{(Eq 16)}$$

The scuffing temperature determined from Forschungsstelle für Zahnräder und Getriebebau (FZG, or Technical Institute for the Study of Gears and Drive Mechanisms) test gears for mineral oils without antiscuff additives or with low concentrations of antiscuff additives may be extended to different gear steels, heat treatments, or surface treatments by introducing an empirical welding factor:

$$T_s = T_{b_{test}} + X_w T_{f_{test}} \qquad \textbf{(Eq 17)}$$

where X_w is the welding factor (see Table 2), $T_{b_{test}}$ is the bulk temperature of the test gears, and $T_{f_{test}}$ is the maximum flash temperature of the test gears.

Table 2 Welding factor for selected gears steels

Material	Welding factor, X_w
Through-hardened steel	1.00
Phosphated steel	1.25
Copper-plated steel	1.50
Nitrided steel	1.50
Carburized steel	
Content of austenite < average	1.15
Content of austenite = average	1.00
Content of austenite > average	0.85
Stainless steel	0.45

Source: Ref 18

Scuffing temperatures for synthetic lubricants typically used with carburized gears in the aerospace industry are shown in Table 3.

Table 3 Mean scuffing temperature for synthetic lubricants typically used for operating carburized gears in aerospace applications

Lubricant	Mean scuffing temperature, T_s	
	°C	**°F**
MIL-L-6081 (grade 1005)	129	264
MIL-L-7808	205	400
MIL-L-23699	220	425
DERD2487	225	440
DERD2497	240	465
DOD-L-85734	260	500
Mobil SHC624	280	540
Dexron II	290	550

Source: Ref 20

For mineral oils with high concentrations of antiscuff additives (for example, hypoid gear oils), research is still needed to determinewhether the scuffing temperature is dependent on the materials and operating conditions. Special attention has to be paid to the correlation between test conditions and actual or design conditions.

Lubricant Selection

The choice of lubricant depends on the type of gearing and enclosure, operating speed and load, ambient temperature, and method of lubricant application. Most gears are lubricated with one of the following types:

- Oil
- Grease
- Adhesive open-gear lubricant
- Solid lubricant

The optimum lubricant for any application is the product that is the least expensive, considering both initial cost and maintenance costs, and meets the requirements.

Oil is the most widely used lubricant because it is readily distributed to gears and bearings and has both good lubricating and cooling properties. In addition, contamination may be readily removed by filtering periodically or draining and replacing the oil. However, it requires an oil-tight enclosure provided with adequate shaft seals.

Synthetic lubricants are used for applications (for example, aircraft gas turbines) where the oil must operate over a wide temperature range and have good oxidation stability at high temperature. Ester and hydrocarbon synthetic lubricants have high viscosity indices, giving them good fluidity or low viscosities at very low temperatures and acceptable viscosities at high temperatures. The volatility of esters is lower than that of mineral oils of the same viscosity, thus reducing oil loss at high temperature. Despite their long service life, the extra cost of synthetic lubricants generally cannot be justified for oil-bath systems unless there are extreme temperatures involved, because the oil must be changed frequently to remove contamination.

Grease is suitable only for low-speed, low-load applications because it does not circulate well, and it is a relatively poor coolant. Grease-lubricated gears are generally boundary lubricated because the grease is either pushed aside or thrown from the gear teeth. Contamination from wear particles or other debris is usually trapped in the grease and requires costly maintenance to eliminate. Grease is often used to avoid leakage from enclosures that are not oil tight. However, if all the factors are considered, it is usually found that an oil lubricant is more economical and reliable than a grease lubricant for gear lubrication.

Open-gear lubricants are viscous adhesive semifluids used on large low-speed open gears, such as those used in iron ore and cement mills, antenna drives, bridge drives, cranes, and so on. Gears in these applications run slowly, and they are therefore boundary lubricated. The lubricant must bond strongly to resist being thrown off the gear teeth. However, the squeezing and sliding action of gear teeth tends to push the lubricant into the roots of the gear teeth where it is relatively ineffective. These lubricants are applied by hand brushing or by automatic systems that deliver an intermittent spray. Some open-gear lubricants are thinned with a quick-evaporating solvent/diluent to make them easier to apply. Open-gear lubricants share the disadvantages of grease lubrication, and they are especially costly (in addition to being messy) to maintain. For these reasons, the trend is away from open gears and toward the use of enclosed oil-lubricated gearboxes whenever possible.

Solid lubricants, usually in the form of bonded dry films, are used when the following conditions are encountered:

- The temperature is too high or too low for an oil or grease
- Leakage cannot be tolerated
- The gears must operate in a vacuum

These lubricants are usually molybdenum disulfide (MoS_2) or graphite in a inorganic binder, which is applied to the gear teeth and cured to form a dry film coating. Polytetrafluoroethylene (PTFE) and tungsten disulfide (WS_2) coatings are also

used. Solid lubricants are expensive to apply and have limited wear lives. However, in many applications, such as spacecraft, they are the only alternative and can provide excellent service.

Oil Lubricant Applications

Of the above-mentioned lubricants, only oil will be discussed in greater detail in this article. Oil should be used as the lubricant unless the operating conditions preclude its use. Generally, the simplest and least expensive lubrication system for gears is a totally enclosed oil bath of mineral oil.

Spur, Helical, and Bevel Gears. The lubrication requirements of spur, helical, straight-bevel, and spiral-bevel gears are essentially the same. For this class of gears, the magnitudes of the loads and sliding speeds are similar, and requirements for viscosity and antiscuff properties are virtually identical. Many industrial spur and helical gear units are lubricated with rust and oxidation inhibited (R & O) mineral oils. The low-viscosity R & O oils, commonly called turbine oils, are used in many high-speed gear units where the gear tooth loads are relatively low. Mineral oils without antiscuff additives are suitable for high-speed lightly loaded gears where the high entraining velocity of the gear teeth develops thick EHD oil films. In these cases, the most important property of the lubricant is viscosity. Antiscuff/EP additives are unnecessary because the gear teeth are separated, eliminating metal-to-metal contact and the scuffing mode of failure. Slower speed gears, especially carburized gears tend to be more heavily loaded. These gears generally require higher viscosity lubricants with antiscuff additives.

Hypoid gears, such as those used for automotive axles, are especially prone to scuffing because they are heavily loaded and have high sliding velocities. For these reasons, hypoid gear oils have the higher concentrations of antiscuff additives.

For critical applications, the contact temperature should be calculated with Blok's equation (see Eq 1) (Ref 15) and compared to the scuffing temperature of the lubricant. This quantitative method is effective for selecting a lubricant with adequate scuffing resistance.

Worm gears have high sliding velocity, which generates significant frictional losses. Fortunately, their tooth loads are relatively light, and they are successfully lubricated with mineral oils that are compounded with lubricity additives. These oils contain 3 to 10% fatty oil or low-acid tallow. The polar molecules of the additive form surface films by physical adsorption or by reaction with the surface oxide to form a metallic soap that acts as a low shear strength film, improving the "lubricity," or friction-reducing property.

Selection of Gear Lubricant Viscosity

The recommendations of AGMA 250.04 (Ref 21) should be followed when selecting lubricants for enclosed gear drives that operate at pitch line velocities $\leq$1500 m/min ($\leq$5000 ft/min). AGMA 421.06 (Ref 22) should be consulted for high-speed drives (>1500 m/min, or 5000 ft/min).

As discussed earlier in the section "Gear Tooth Failure Modes" in this article, viscosity is one of the most important lubricant properties, and the higher the viscosity, the greater the protection against the various gear tooth failures. However, the viscosity must be limited to avoid excessive heat generation and power loss from churning and shearing of the lubricant by high-speed gears or bearings. The operating temperature of the gear drive determines the operating viscosity of the lubricant. If the lubricant is too viscous, excessive heat is generated. The heat raises the lubricant temperature and reduces its viscosity, reaching a point of diminishing returns where increasing the starting viscosity of the lubricant leads to a higher operating temperature and a higher oxidation rate, without a significant gain in operating viscosity.

Gear drives operating in cold climates must have a lubricant that circulates freely and does not cause high starting torques. A candidate gear lubricant should have a pour point at least 5 °C (9 °F) lower than the expected minimum ambient start-up temperature. Typical pour points for mineral gear oils are -7 °C (20 °F), whereas synthetic gear lubricants have significantly lower pour points of about -40 °C (-40 °F). Pour point depressants are used to tailor pour points of mineral lubricants for automotive hypoid gears to be as low as -40 °C (-40 °F).

The pitch line speed of the gears is a good index of the required viscosity. An empirical equation for determining required viscosity is:

$$\nu_{40} = \frac{7000}{(V)^{0.5}} \quad \textbf{(Eq 18)}$$

where ν_{40} is the lubricant kinematic viscosity at 40 °C (105 °F) (in cSt) and V is the operating pitch line velocity (in ft/min) given by:

$$V = 0.262\, d \cdot n \quad \textbf{(Eq 19)}$$

where d is the operating pitch diameter of pinion (in inches) and n is the pinion speed (rev/min).

Caution must be used when using AGMA recommendations for viscosity. The author knows of an application where two gear drives were considered to be high speed. The pinion speed was 3625 rev/min, qualifying the gear units as high-speed gear drives per AGMA 421.06. The gear drives were supplied with oil having the recommended viscosity (per AGMA 421.06) of ISO 68. However, because the pinion was relatively small, its pitch line velocity was only 1000 m/min (3000 ft/min). This qualifies the gear drives as slow speed per AGMA 250.04, which recommends a viscosity of ISO 150. Both gear drives failed within weeks of start-up because of pitting fatigue. The empirical equation (Eq 18) for this application gives:

$$\nu_{40} = \frac{7000}{(3000)^{0.5}} = 128 \text{ cSt} \quad \textbf{(Eq 20)}$$

This indicates that the viscosity per AGMA 421.06 (68 cSt, or 6.8×10^{-5} m^2/s) is much too low, and the viscosity per AGMA 250.04 (150 cSt, or 1.5×10^{-4} m^2/s) is appropriate. Hence, definitions of high-speed versus slow-speed gear drives must be carefully considered, and pitch line velocity is generally a better index than shaft speed. The gear drives were rebuilt with new gearsets and the ISO VG 68 oil was replaced with ISO VG 150. The gear drives now operate without overheating, and the pitting has been eliminated.

For critical applications, the specific film thickness, λ, should be calculated with Dowson and Higginson's equation (see Eq 8) (Ref 16). The specific film thickness is a useful measure of the lubrication regime. It can be used with Fig. 6 as an approximate guide to the probability of wear-related surface distress. Figure 6 is based on the data of Wellauer and Holloway (Ref 23), which were obtained from several hundred laboratory tests and field applications of gear drives.

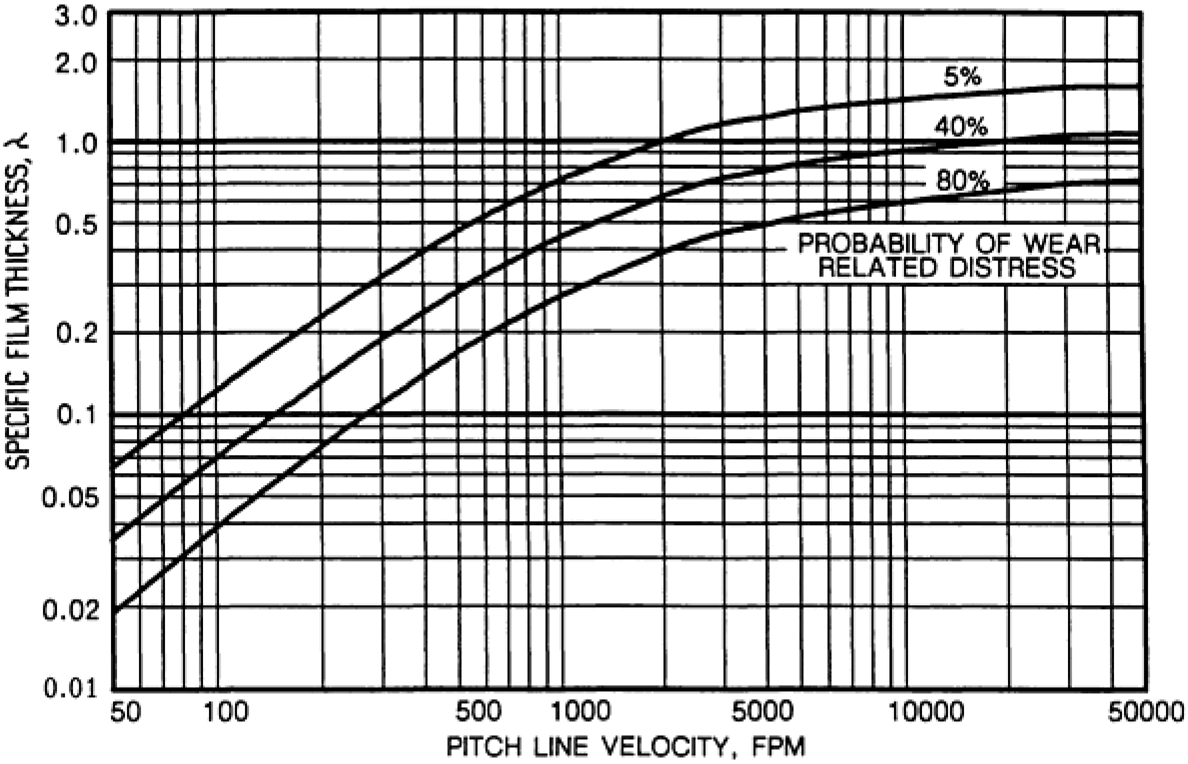


Fig. 6 Probability of wear distress as a function of specific film thickness and pitch line velocity (in ft/min). Source: Ref 18

Application of Gear Lubricants

The method of applying the lubricant to the gear teeth depends primarily on the pitch line velocity.

Splash lubrication systems are the simplest, but they are limited to a pitch line velocity of ~1000 m/min (~3000 ft/min). The gears should dip into the oil bath for about twice the tooth depth to provide adequate splash for pinions and bearings and to reduce losses due to churning. The gear housing should have troughs to capture the oil flowing down the housing walls, channeling it to the bearings.

The range of splash lubrication can be extended to ~1500 m/min (~5000 ft/min) by using baffles and oil pans to reduce churning. However, at velocities >1000 m/min (>3000 ft/min), providing auxiliary cooling with fans and improving heat transfer by adding fins to the housing are usually necessary.

Pressure-Fed Systems. Above 1500 m/min (5000 ft/min), most gears are lubricated by a pressure-fed system. For gearboxes with anti-friction bearings, spraying the oil at the gear mesh only and relying on splash to lubricate the bearings is permissible up to a maximum pitch line velocity of 2100 m/min (7000 ft/min). Above this speed and for gear drives with journal bearings, both the gears and bearings should be pressure-fed.

The oil jets should be placed on the incoming side of the gear mesh for pitch line velocities ≤2600 m/min (≤8000 ft/min). Above 2600 m/min (8000 ft/min), more oil is needed for cooling than for lubricating, and the oil flow removes heat best by being directed at the outgoing side of the gear mesh where the oil jets can strike the hot drive-side of the gear teeth.

For very high-speed gears (Ref 19) (>5300 m/min, or 16,000 ft/min), there is a danger that the amount of oil carried to the incoming side of the gear mesh may be inadequate, and it is prudent to add a supplementary flow at the incoming side of the gear mesh. Generally, about two-thirds of the oil flow should be supplied to the outgoing side of the mesh for cooling,

and one-third of the flow directed at the incoming side for lubrication. The placement of the oil jets is a crucial factor when pitch line velocities are >6100 m/min (>20,000 ft/min). At speeds this high, experiments are required to find the optimum number and location for the oil jets.

In pressure-fed systems, the following parameters must be considered to ensure adequate lubrication and cooling of the gear mesh:

- Quantity of flow
- Jet size
- Feed pressure
- Number of jets

There are general guidelines, based on experience and experimentation, for specifying these parameters, but each application must be evaluated independently based on its particular operating conditions and requirements.

An empirical equation used to calculate the quantity of oil flow, *q,* is:

$$q = P/c \qquad \textbf{(Eq 21)}$$

where q is the oil flow rate (in gal/min), P is the transmitted power (in hp), and c is taken from Table 4.

Table 4 Recommended values of constant *c* based on oil flow and gear mesh specifications

c hp/(gal/min)	Flow conditions	Applications
200	Copious	General industrial
400	Adequate	Typical aviation
800	Lean	Lightweight, high-efficiency aviation
1000	Starved	Only for unusual conditions

For a typical industrial application transmitting 150 kW (200 hp), where weight is not critical, the designer might choose a value of c = 200 hp/(gal/min), resulting in a copious flow of 4 L/min (1 gal/min). On the other hand, for a high-efficiency aviation application transmitting 150 kW (200 hp), where weight is critical, a value of c = 800 hp/(gal/min) might be chosen, resulting in a lean flow of 1 L/min (0.25 gal/min). Some applications may require different flow rates than those given by Table 4. For instance, wide-face, high-speed gearing may require a higher flow rate to ensure uniform cooling and full-face coverage.

The proper jet size, feed pressure, and number of jets must be determined to maintain the proper flow rate, jet velocity, and full-face coverage.

Jet Size. The diameter of a jet can be calculated for a given flow rate and pressure based on the viscosity of the oil at the operating temperature (Ref 24). There are practical limitations on jet size, and the minimum recommended size is 0.8 mm (0.03 in.). If a jet smaller than this is used, contaminants in the oil may clog it. Typical jet diameters range from 0.8 to 3.0 mm (0.03 to 0.12 in.).

Feed Pressure. The feed pressure determines the jet velocity, which in turn determines the amount of oil that penetrates the gear mesh. Typical feed pressures range from 20 to 100 psig. Industrial application feed pressures are typically 30 psig, and high-speed aerospace applications are typically 100 psig. In general, the higher the pressure, the greater the cooling (Ref 25), but the higher the pressure, the smaller the jet diameter. Therefore, pressure is limited by the minimum recommended jet diameter of 0.8 mm (0.03 in.).

Jet Quantity. The number of jets should be sufficient to provide complete lubrication coverage of the face width. More than one jet for each gear mesh is advisable because of the possibility of clogging. The upper limit on the number of jets

is determined by the flow rate and jet diameter; too many jets for a given flow rate will result in a jet diameter less than the minimum recommended.

Example: Friction and Wear in a 24-Unit Speed-Increaser Gearbox.

In an industrial application, 24 speed-increaser gearboxes were used to transmit 258 kW (346 hp) and increase speed from 55 to 375 rev/min. The gears were parallel shaft, single helical, carburized, and ground. The splash lubrication system used a mineral oil without antiscuff additives with ISO 100 viscosity. After about 250 h of operation, two gearboxes failed by bending fatigue. The gear tooth profiles were so badly worn that determining the primary failure mode was impossible. Three other gearboxes with less service were selected for inspection. One had logged 15 h, and the other two had operated for 65 h each. Upon disassembly, no broken teeth were found, but all three gearboxes had scuffed gear teeth. The primary failure mode was scuffing, and the earlier bending fatigue failures were caused by dynamic loads generated by the worn gear teeth. Subsequent inspection of the remaining gearboxes revealed that all had scuffing damage, which probably had occurred immediately upon start-up because the loads were not reduced during run-in.

Fortunately, a prototype gearbox had been run at one-half load for about 50 h. When these gears were inspected, no signs of distress were seen on any of the gear teeth. The tooth profiles were smooth with surface roughness. R_q, estimated to be 0.5 μm (20 μin.), and the contact pattern indicated 100% face contact. This gearbox was reassembled and run under one-half load until its oil sump temperature reached equilibrium at 95 °C (200 °F). For this application, the ambient temperature was in the range of 10 to 50 °C (50 to 125 °F). The center distance of the gears was 405 mm (16 in.) and the pitch line velocity was 120 m/min (400 ft/min). Referring to AGMA 250.04 (Ref 21), the recommended viscosity for these conditions is ISO 150 or ISO 220.

Using the empirical equation (Eq 18), we obtain

$$\nu_{40} = \frac{7000}{(400)^{0.5}} = 350 \text{ cSt} \qquad \textbf{(Eq 22)}$$

Hence, the empirical equation recommends a viscosity close to ISO 320. It is apparent that the viscosity that was originally supplied (ISO VG 100) was too low.

The EHD film thickness was calculated with a special computer program (Ref 26). The gear bulk temperature was assumed to be 110 °C (230 °F), which is 17 °C (30 °F) hotter than the measured oil sump temperature. The following data for the ISO VG 100 lubricant was obtained from Fig. 3 and 4: μ_0 = 6.6 cP = 0.96 × 10^{-6} reyns and α= 1.02 × 10^{-4} in.^{2}/lbf.

Figure 7 shows a plot of the film thickness versus position on the pinion tooth. The minimum film thickness occurs low on the pinion tooth near the lowest point of single-tooth contact (LPSTC), where h_{min} is 0.053 μm (2.1 μin.). The specific film thickness, based on R_q of 0.5 μm (20 μin.) for both profiles, is λ= 0.073. Figure 6 shows that the gears operate in the boundary lubrication regime. The program predicts that the probability of wear is >95%.

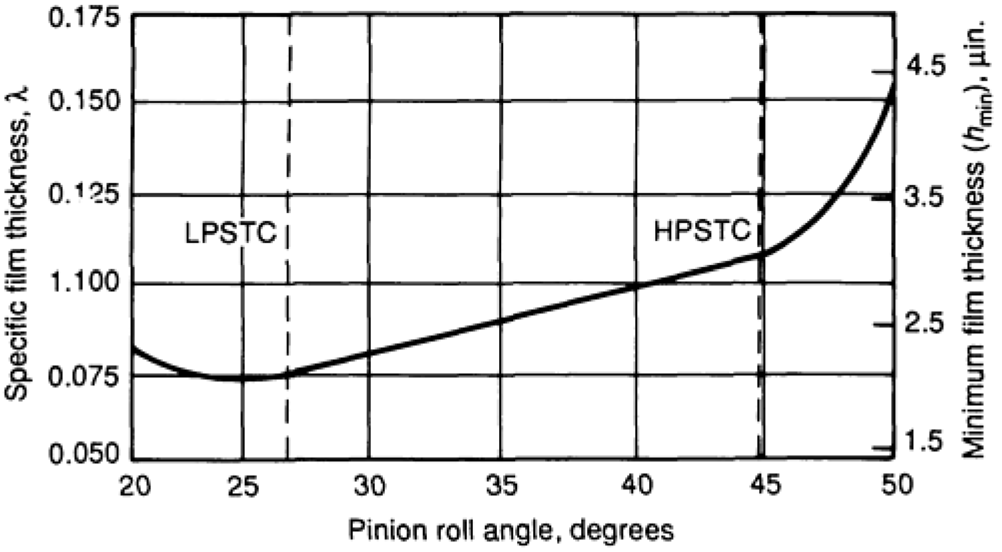


Fig. 7 Plot of film thickness versus pinion roll angle for gear tooth geometry of a scuffed gearset. Minimum specific film thickness, λ_{min}, 0.073; probability of wear, >95%

The contact temperature was also calculated with the program. The scuffing temperature for the ISO VG 100 lubricant was calculated with the equation for nonantiscuff mineral oils:

$$T_s = 146 + 59 \ln (100) = 418 \text{ °F} \quad \textbf{(Eq 23)}$$

Figure 8 shows a plot of the contact temperature versus position on the pinion tooth. The maximum contact temperature occurs high on the pinion tooth near the highest point of single-tooth contact (HPSTC), where T_c is 226 °C (439 °F). The program predicts that the probability of scuffing is 63%. This is considered to be a high risk of scuffing. The relatively high-temperature peak near the tip of the pinion tooth was caused by the geometry of the gears.

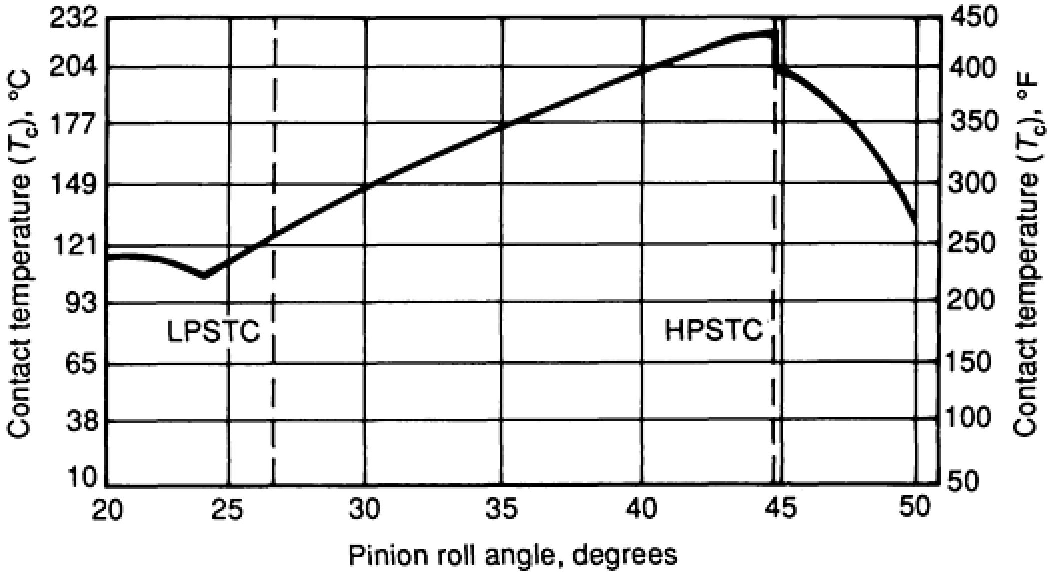


Fig. 8 Plot of contact temperature versus pinion roll angle for gear tooth geometry of scuffed gearset. Maximum T_f, 226 °C (439 °F); scuffing probability, 63%

The designer selected a long-addendum tooth for the pinion. Long-addendum pinions perform well in speed reducers, where they increase the amount of recess action and decrease the amount of approach action of the gear mesh. Because recess action is much smoother than approach action, long-addendum pinions give speed smooth meshing characteristics. When operated as a speed increaser, however, the approach and recess portions of the gear mesh reverse, making a long-addendum pinion rough running and vulnerable to scuffing.

To explore the possibilities for reducing the scuffing risk, new gear tooth geometry was proposed with the pinion and gear addenda designed to minimize the flash temperature rise. The new gearset, analyzed with the program, assumed the lubricant was a mineral oil with antiscuff additives, with a viscosity of ISO 220, and with the following properties: μ_0 = 10 cP = 1.45 × 10^{-6} reyns, α= 1.09 × 10^{-4} in.^{2}/lbf, and

$$T_s = 245 + 59 \ln (220) = 563 \text{ °F} \quad \textbf{(Eq 24)}$$

Figure 9 shows that the film thickness increases to h_{min} of 0.068 μm (2.7 μin.), and the specific film thickness increases to λ= 0.097. Figure 6 shows that the gears still operate in the boundary lubrication regime; however, the probability of wear is reduced to 94%. Figure 10 shows that the optimized gear geometry reduced the maximum contact temperature to a T_c of 150 °C (302 °F). The combination of reduced contact temperature and the increased scuffing resistance provided by the higher viscosity mineral oil with antiscuff additives reduces the scuffing probability to <5%.

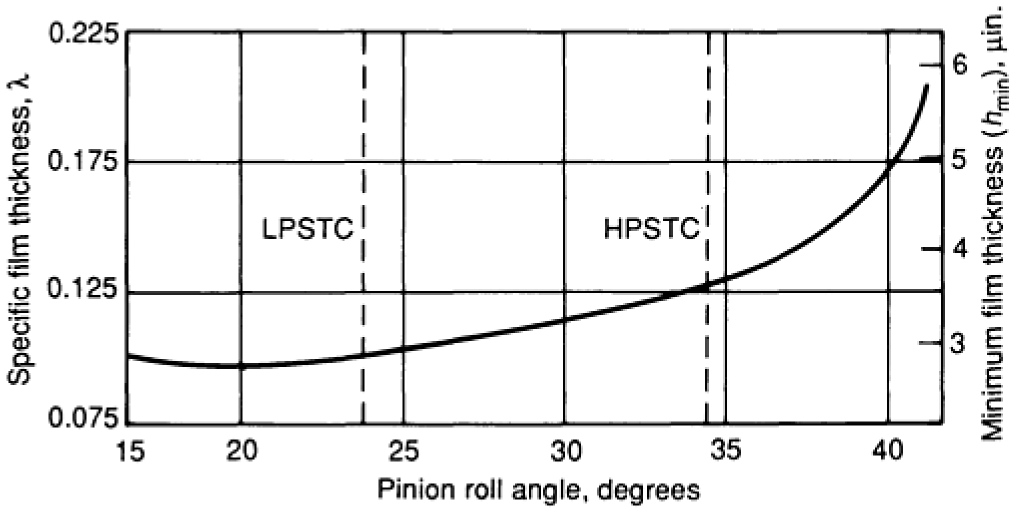


Fig. 9 Plot of film thickness versus pinion roll angle for gear tooth geometry that was optimized for maximum scuffing resistance. λ_{min}, 0.097; probability of wear, 94%

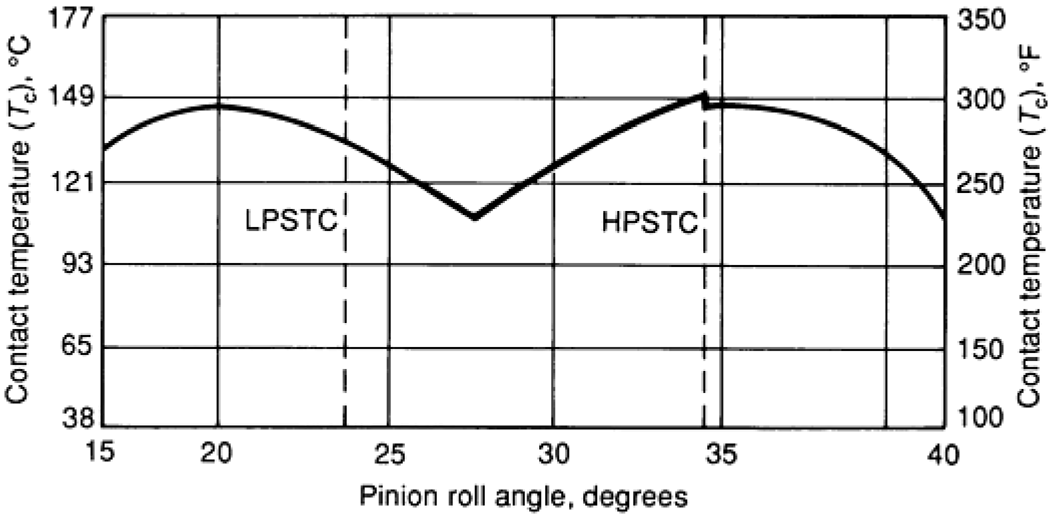


Fig. 10 Plot of contact temperature versus pinion roll angle for gear tooth geometry that was optimized for maximum scuffing resistance. Maximum T_f, 150 °C (302 °F); scuffing probability, <5%

Typical of may gear failures, this example shows that several factors contributed to the failures:

- The lubricant viscosity was too low
- No antiscuff additives were used
- A gearbox designed as a speed reducer was used as a speed increaser
- The gear teeth were not provided with a coating or plating to ease running-in
- The gears were not run-in properly under reduced loads

Gear failures, as exemplified by this case history, can be avoided if designers and operators recognize that the lubricant is an importantcomponent of a gearbox and appreciate that the tribology of gearing requires the consideration and control of many interrelated factors.

Nomenclature

Table 5 lists the parameters used in this article and the units required to calculate Eq 1 through 24.

Table 5 Nomenclature used in friction and wear of gears

Symbol	Description	Units
B_M	Thermal contact coefficient	lbf/(in. · s$^{0.5}$ · °F)
b_H	Semiwidth of Hertzian contact band	in.
c	Constant (see Table 4.)	hp/(gal/min)
c_M	Specific heat per unit mass	lbf · in./[lb · °F]
d	Operating pitch diameter of pinion	in.
E_1	Modulus of elasticity of pinion	lbf/in.^{2}
E_2	Modulus of elasticity of gear	lbf/in.^{2}
E_r	Reduced modulus of elasticity	lbf/in.^{2}
h_{min}	Minimum film thickness	in.
L_{min}	Minimum contact length	in.
n	Pinion speed	rev/min
P	Transmitted power	hp
q	Oil flow rate	gal/min
S	Average root-mean-square (rms) surface roughness	μin.

T_b	Bulk temperature	°F
$T_{b_{test}}$	Bulk temperature of test gears	°F
T_c	Contact temperature	°F
T_f	Flash temperature	°F
$T_{f_{test}}$	Maximum flash temperature of test gears	°F
T_s	Scuffing temperature	°F
V	Operating pitch line velocity	ft/min
V_e	Entraining velocity	in./s
V_{r1}	Rolling velocity of pinion	in./s
V_{r2}	Rolling velocity of gear	in./s
W_{Nr}	Normal operating load	lbf
w_{Nr}	Normal unit load	lbf/in.
X_W	Welding factor	. . .
X_Γ	Load sharing factor	. . .
α	Pressure-viscosity coefficient	in.^{2}/lbf
λ	Specific film thickness	. . .
λ_{min}	Minimum specific film thickness	. . .
λ_M	Heat conductivity	lbf/[s · °F]
μ_m	Mean coefficient of friction	. . .
μ_0	Absolute viscosity	reyns (lbf · s/in.^{2})
v_1	Poisson's ratio of pinion	. . .
v_2	Poisson's ratio of gear	. . .
v_{40}	Kinematic viscosity at 40 °C (105 °F)	cSt
ρ_1	Transverse radius of curvature of pinion	in.
ρ_2	Transverse radius of curvature of gear	in.
ρ_M	Density	lb/in.^{3}
ρ_n	Normal relative radius of curvature	in.
σ	Root-mean-square composite surface roughness	μin.
σ_1	Root-mean-square surface roughness of pinion	μin.
σ_2	Root-mean-square surface roughness of gear	μin.
ψ_b	Base helix angle	degrees
ω_1	Angular velocity of pinion	rad/s
ω_2	Angular velocity of gear rad/s	rad/s

References

1. "Nomenclature of Gear Tooth Failure Modes," ANSI/AGMA 110.04, American Gear Manufacturers Association, 1980
2. E.E. Shipley, Gear Failures, *Mach. Des.,* 7 Dec 1967, p 152-162
3. D.W. Dudley, Gear Wear, *Wear Control Handbook,* American Society of Mechanical Engineers
4. P.M. Ku, Gear Failure Modes--Importance of Lubrication and Mechanics, *ASLE Trans.,* Vol 19 (No. 3), 1975, p 239-249
5. D.J. Wulpi, *Understanding How Components Fail,* American Society for Metals, 1985
6. Failures of Gears, *Failure Analysis and Prevention,* Vol 10, 8th ed., *Metals Handbook,* American Society for Metals, p 507-524
7. D. Godfrey, Recognition and Solution of Some Common Wear Problems Related to Lubrication and Hydraulic Fluids, *Lubr. Eng.,* Feb 1987, p 111-114
8. W.E. Littman, The Mechanism of Contact Fatigue, *Interdisciplinary Approach to the Lubrication of Concentrated Contacts,* SP-237, National Aeronautics and Space Administration, 1970, p 309-377
9. T. Ueno *et al.,* "Surface Durability of Case-Carburized Gears--On a Phenomenon of Grey Staining of Tooth Surface," Paper No. 80-C2/DET-27, American Society of Mechanical Engineers, 1980, p 1-8
10. H. Winter and T. Weiss, "Some Factors Influencing the Pitting, Micropitting (Frosted Areas) and Slow Speed Wear of Surface-Hardened Gears," Paper No. 80-C2/DET-89, American Society of Mechanical

Engineers, 1980, p 1-7

11. E.E. Shipley, "Failure Analysis of Coarse-Pitch, Hardened, and Ground Gears," Paper No. P229.26, American Gear Manufacturers Association, 1982, p 1-24
12. S. Tanaka *et al.*, "Appreciable Increases in Surface Durability of Gear Pairs with Mirror-Like Finish," Paper No. 84-DET-223, American Society of Mechanical Engineers, 1984, p 1-8
13. A. Milburn, R. Errichello, and D. Godfrey, "Polishing Wear," Paper No. 90 FTM 5, American Gear Manufacturers Association, 1990, p 1-13
14. J.H. Admas and D. Godfrey, Borate Gear Lubricant-EP Film Analysis and Performance, *Lubr. Eng.*, Vol 37 (No. 1), Jan 1981, p 16-21
15. H. Blok, "Les Temperatures de Surface dans les Conditions de Graissage sons Pression Extreme," Second World Petroleum Congress (Paris), June 1937
16. H. Blok, The Postulate about the Constancy of Scoring Temperatures, *Interdisciplinary Approach to the Lubrication of Concentrated Contacts*, SP-237, National Aeronautics and Space Administration, 1970, p 153-248
17. D. Dowson, Elastohydrodynamics, Paper No. 10, *Proc. Inst. Mech. Eng.*, Vol 182, PT3A, 1967, p 151-167
18. "Fundamental Rating Factors and Calculation Methods for Involute Spur and Helical Gear Teeth," 2001-B88, American Gear Manufacturers Association, 1988
19. M. Akazawa, T. Tejima, and T. Narita, "Full Scale Test of High Speed, High Powered Gear Unit--Helical Gears of 25,000 PS at 200 m/s PLV," Paper No. 80-C2/DET-4, American Society of Mechanical Engineers, 1980
20. R.J. Drago, "Comparative Load Capacity Evaluation of CBN-Finished Gears," Paper No. 88 FTM 8, American Gear Manufacturers Association, Oct 1988
21. "AGMA Standard Specification--Lubrication of Industrial Enclosed Gear Drives," No. 250.04, American Gear Manufacturers Association, Sept 1981
22. "Practice for High Speed Helical and Herringbone Gear Units," No. 421.06, American Gear Manufacturers Association, Jan 1969
23. E.J. Wellauer and G.A. Holloway, Application of EHD Oil Film Theory to Industrial Gear Drives, *J. Eng. Ind. (Trans. ASME)*, Vol 98, Series B (No. 2), May 1976, p 626-634
24. R.J. Drago, *Fundamentals of Gear Design*, Butterworths, 1988
25. L. Akin and D. Townsend, "Study of Lubricant Jet Flow Phenomena in Spur Gears," TMX-71572, National Aeronautics and Space Administration, Oct 1974
26. SCORING+, computer program, GEARTECH Software, Inc., Copyright 1985-1992

Friction and Wear of Seals

James F. Dray, Mechanical Technology Inc.

Introduction

SEALS are the components in a mechanical system that act as the pressure boundary between other components, such as a piston and cylinder in a hydraulic actuator. A region of high pressure usually exists on one side of the seal, versus low pressure on the other.

In some cases, the role of the seal is to prevent mixing of two incompatible fluids, more so than it is to hold a pressure boundary. The other components can be either stationary or in motion. A diaphragm is an example of a seal that effectively prevents mixing. The fluids being sealed can be either liquids or gases.

Types of Seals

Seals are usually classified as being either static or dynamic, which refers to the motion the seal experiences rather than the presence or absence of pressure fluctuations. A dynamic seal has a surface that is in motion (usually sliding) relative to the contacting surface. This type of seal can experience either rotary or reciprocating motion or a combination of both.

Static seals are available in many variations. The most frequently used static seals are probably O-rings, gaskets, and packing. O-rings (Fig. 1) are also occasionally used as dynamic seals, and are effective as seals when used within their recommended range.

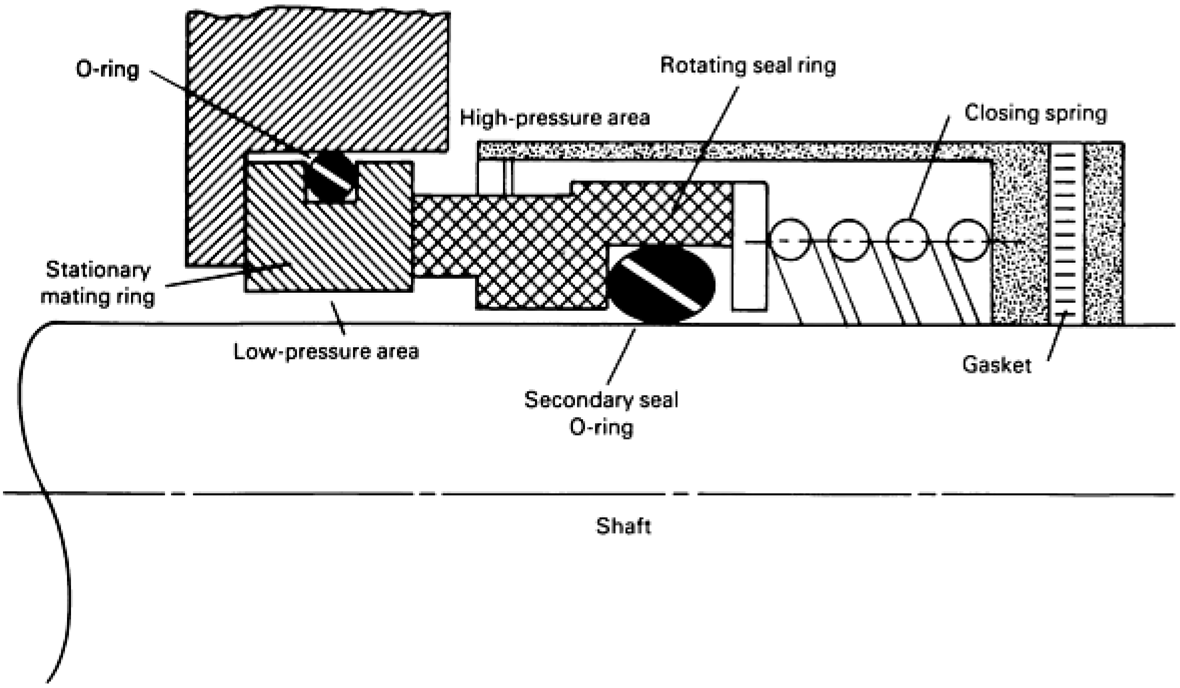


Fig. 1 Mechanical face seal

O-rings are made from a variety of polymer materials, including natural, neoprene, and nitrile rubbers. They are generally limited to a temperature range from 0 to 93 °C (32 to 200 °F), because of their material composition. Chemical compatibility between the seal material and the fluid being sealed is an important design requirement. The surface finishes of the groove containing the O-ring and the mating surface are also significant in determining seal life and effectiveness.

Another design requirement is the amount of compression. The catalogs of O-ring manufacturers provide definitive tables of groove size versus O-ring size and can determine compression, which is generally from 10 to 15%.

Gaskets (Fig. 1) are used as seals between static machine joints and are the least expensive, most effective choice for many applications. They are available in a variety of materials, ranging from cork to metals such as copper. Some, such as room-temperature vulcanizing (RTV) silicone rubber, can be formed in place. Many contain asbestos, which is being phased out because of environmental and health risks.

The material selection and joint design of gaskets are driven by operating temperature range, chemical compatibility, and differential pressure. Typical applications are an engine head to block, steam turbine upper to lower housing, and bearing pillow block housing joints.

Packing is also used for these static applications in place of a gasket because of ready availability, low cost, and simplicity of design and applicability. Another advantage is that it allows intermittent operation of the joint, as is required for ship rudders and submarine periscopes. Packing types have evolved from braided flax, filled with tallow as a lubricant, to modern materials, such as braided polytetrafluoroethylene (PTFE) fiber, impregnated with petrolatum.

Packing is usually square in cross section, and the joint or stuffing box design is controlled by the same parameters that control a gasket (Fig. 2). It can be staged for high-pressure applications, and its compression is controlled by tightening the gland bolts.

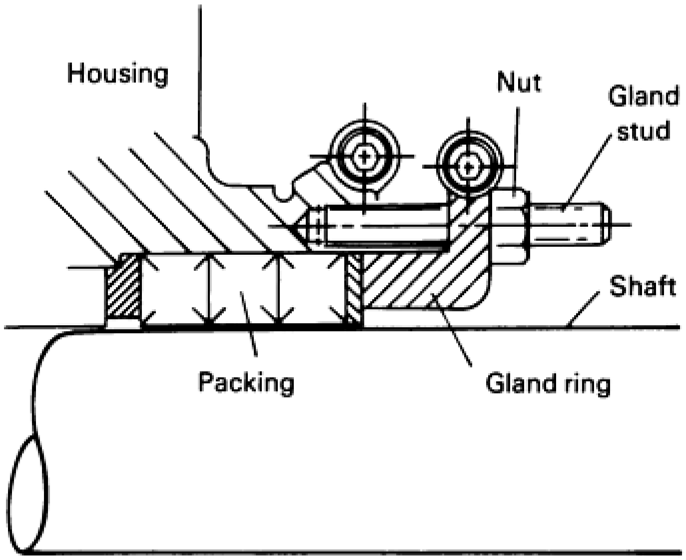


Fig. 2 Packing gland

Dynamic seals can experience either rotary or reciprocating motion or a combination of both. There are a number of different configurations, including lip, face, cap, quad ring, O-ring, and packing, as well as hybrid combinations of these.

Lip seals are widely used to seal low-pressure differentials up to about 137.8 kPa (20 psi) and are limited to relatively small amounts of radial motion between the journal and seal (that is, less than 0.508 mm, or 20 mils). The lips are loaded against the shaft by the pressure differential, or by finger springs molded into the lips, or by a garter spring wrapped around the lip (Fig. 3). When relative velocities and lip loads are high, a considerable amount of heat is generated and the lip may either groove or cut the shaft unless some form of lubrication is provided (Ref 1). Materials such as nitrile rubber must be lubricated with oil or grease, whereas self-lubricating materials such as PTFE provide their own lubrication. Many of these materials also have their lubricity enhanced by the addition of a dispersion of such natural lubricants as graphite or molydisulfide.

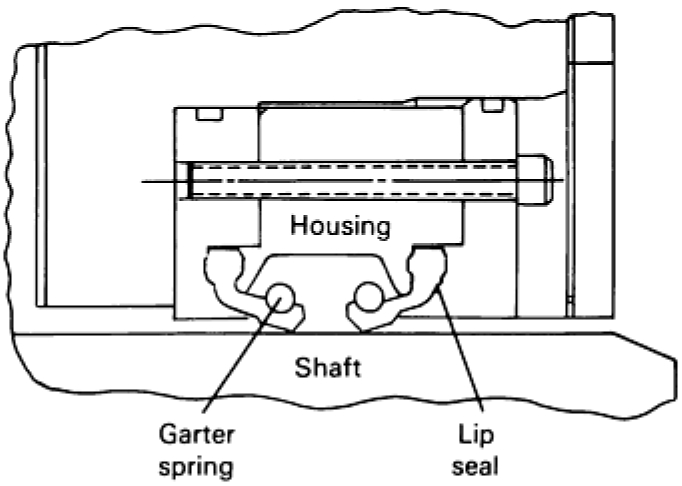


Fig. 3 Lip seal

Face seals are usually reserved for higher-pressure applications (Ref 2). They are used in such diverse sizes and applications as surface ship and submarine main propulsion shafts and high-pressure pump seals (Ref 3). The three elements (Fig. 1) are the mating ring, which in this case is stationary; the seal ring, which rotates with the shaft and floats axially to allow for axial shaft motion and is held against the mating ring by hydrostatic pressure, compression springs, or a combination of both; and an O-ring, which acts as a secondary seal.

Pressure differentials and face loadings (contact pressure) are often high, resulting in considerable heat generation and possibly high wear (Ref 4). At the low end of the load spectrum, elastomers such as nitrile rubber are satisfactory, but at the high end, ceramics, cermets, and composites are frequently necessary to minimize wear. Interface lubrication conditions are often complex, ranging from boundary conditions, to mixed conditions, to the hydrodynamic regime.

Hydrostatic face seals have been highly successful in those applications in which modest amounts of leakage can be tolerated in favor of a markedly reduced wear rate. Spiral grooved seals are a variation in which pumping grooves in one of the faces create a pressure opposing the differential pressure being sealed (Ref 5, 6). They have performed effectively in both liquids and gases, but depend on rotation to generate the seal pressure (Fig. 4).

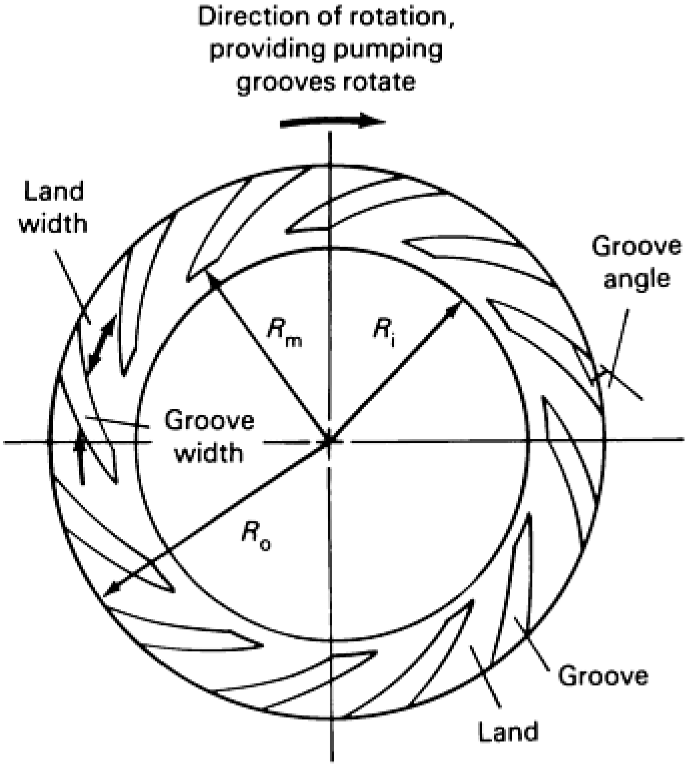


Fig. 4 Spiral groove seal

Piston ring seals are in the reciprocating category and deserve special mention. Engine piston ring seals are always metal, because of the high combustion temperatures, and are split to allow installation over the piston. There are several of these ring seals in series; the upper rings seal the air-fuel mixture and the lower rings seal the lubricating oil splashed on the cylinder walls. Although the rings have a small gap when installed, they provide effective sealing, because of the short duration of the high-combusting gas pressure, until an exhaust valve or port opens. The lubricating oil splashed on the cylinder walls is intended to provide a lubricating film for the piston and rings.

Gas-turbine seals pose a different set of problems because of high operating temperatures and speeds and the low-viscosity fluid being sealed (Ref 7, 8, 9). Compressors have similar rings from a functional viewpoint, but because temperatures can be limited to lower values, the use of polymer rings is increasing. They can be made self-lubricating and can therefore limit the dangers of an explosion (such as dieseling, which is due to the presence of oil) during compression, especially in air compressors. Hydraulic actuators and pumps use similar rings or seals to contain the high-pressure fluid

and are made from a mixture of metal and polymeric materials, depending on the pressure being sealed and the fluid properties.

Labyrinth seals, as their name implies, provide their sealing action via short and segmented tortuous paths (Fig. 5). These seals are ineffective at low speeds because they depend on rotary motion to generate a higher-pressure hydrodynamic fluid film across the close-clearance sections in order to provide the sealing action. They are effective in sealing low pressure differentials (that is, 138 kPa, or 20 psi, and less).

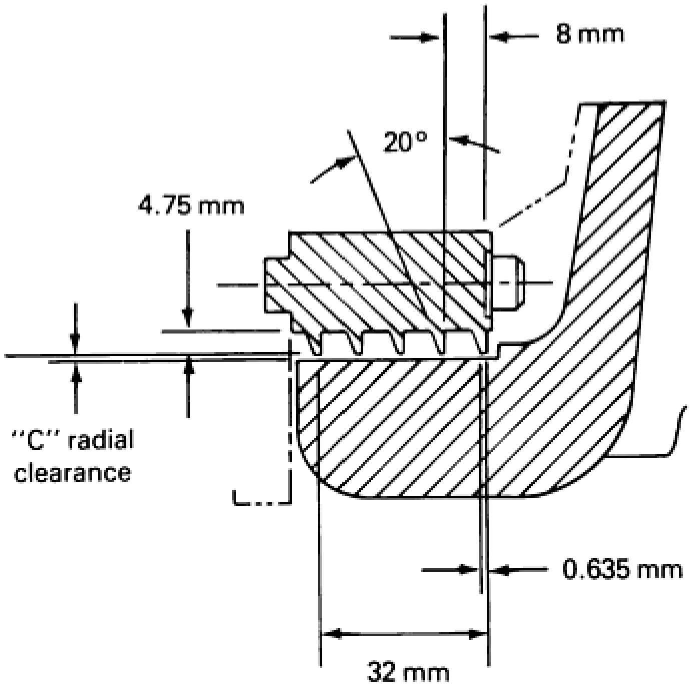


Fig. 5 Labyrinth seal

Hybrid seals that consist of packing integrated with a face seal, for example, can often provide protection in case of primary seal failure in those situations where such a failure, with significantly increased leakage, flooding, and consequent downtime, is intolerable. The packing is activated by fluid pressure leaking by the primary seal during the failure process. Another popular hybrid seal is a labyrinth in tandem with a close-clearance (capillary) floating ring seal. The labyrinth provides effective sealing for low-differential pressures at higher shaft rotational speeds, whereas the floating ring is more effective at lower speeds.

Seal materials include selections from almost all known classes of materials. Magnetic seals use the fluids themselves as the sealant (Ref 10.) Table 1 provides a partial list by application. Generally, those materials that have good tribological properties make good seal materials. The material should have a low coefficient of friction, which means a low shear strength at operating conditions. It also should be readily formable, and it is this property that makes polymers so popular today. Properties of some of the more popular hard face materials used for mechanical face seals are compiled in Table 2, whereas properties of soft face materials are described in Table 3.

Table 1 Seal specifications based on application

Seal application	Type seal	Seal material	Pressure range		Temperature range	
			MPa	**ksi**	**°C**	**°F**
Hydraulic cylinder flange	O-ring	Natural neoprene, nitrile rubber	0-62	0-9	0-95	32-200
	Gasket	Rubber				
Actuator rod	O-ring, quad ring	Neoprene, nitrile rubber	0-62	0-9	0-95	32-200
Engine/turbine manifold	Gasket	Copper, aluminum, asbestos	0-69	0-10	>540	>1000
Main shaft	Packing	Flax, asbestos, Teflon	0-7	(0-1)	0-95	32-200
Bearing housing	Lip	Nitrile rubber, Teflon	0-0.138	0-0.02	0-95	32-200
Wheel bearing	Lip	Synthetic rubber	0-0.034	0-0.005	0-150	32-300
High-pressure pump	Face	Carbon-graphite, silicon carbide	<34.5	<5	<260	<500
Bearing housing	Labyrinth	Aluminum, brass, bronze	<0.138	<0.02	<95	<200
Engine piston ring	Split ring	Chrome-plated cast iron	<20.7	<3	<260	<500
Thrust bearing housing	Floating ring	Bronze	<0.1	<0.025	<95	<200
Gas Turbine	Spiral groove	Bronze	<4.14	<0.6	<650	<1200

Table 2 Properties of hard mating face materials

Property	Product and manufacturer								
	Stellite 3 (sand cast), Haynes	**Hexoloy KT SiC, Carborundum**	**Hexoloy SA or alpha SiC, Carborundum**	**SiC Refel, BNFL**	**Composition B, DTNSRDC**	**Alundum HS (Al_2O_3), Norton**	**High-alumina ceramics (Al_2O_3)**		
							85%	**95%**	**99%**
Composition	. . .	8-13 wt% free Si	. . .	8-12 wt% free Si	. . .	. . .	. . .	. . .	. . .
Apparent density, g/cm^3	8.3	3.09	3.10	3.10	7.9	3.3	3.45	3.65	3.85
Apparent porosity, %	Nil	Nil	2	Nil	Nil	Nil	Nil	Nil	Nil
Specific heat, J/kg · K (Btu/lb · °F)	3.77 (0.9)	0.8 (0.2)	0.67 (0.16)	1.94 (0.5)	0.59 (0.14)	1.63 (0.38)	1.45 (0.35)	1.75 (0.42)	1.95 (0.46)
Thermal conductivity (room temperature), W/m · K (Btu/ft · h · °F)	51.0 (30)	129.7 (75)	125.6 (73)	100.0 (58)	9.7 (5.6)	150.4 (87)	8.5 (5)	12.1 (7)	14.5 (8.4)
Mean coefficient of thermal expansion, 10^{-6}/K	14.0	5.04	4.02	4.32	9.1	7.2	6.7	7.2	7.7
Hardness (Rockwell, Vickers, or Knoop)	55 HRC	2500	2800	2500 HV	43 HRC	2000	1450	1750	2100
Compressive strength, MPa (ksi)	2137 (310)	1000 (145)	3900 (565)	3505 (508)	. . .	2804 (407)	1753 (254)	2103 (305)	2804 (407)
Flexural strength, MPa (ksi)	. . .	383 (55)	550 (80)	. . .	. . .	350 (50)	280 (40)	350 (50)	350-490 (50-71)
Ultimate tensile strength, MPa (ksi)	385 (55)	105 (15)	. . .	315 (45)	524 (75)	210 (30)	140 (20)	210 (30)	273 (40)
Yield strength, MPa (ksi)	. . .	. . .	. . .	. . .	371 (55)	259 (38)	. . .	. . .	. . .
Compressive modulus, GPa (10^6 psi)	. . .	. . .	3955 (575)	. . .	. . .	. . .	. . .	. . .	. . .
Elastic modulus, GPa (10^6 psi)	231 (34)	380 (55)	410 (59)	415 (60)	255 (37)	280 (41)	224 (33)	280 (41)	350 (51)

Table 3 Properties of soft face materials

Property	Product and manufacturer						
	Carbon graphites					**P-658RC, Pure Carbon**	**P-692, Pure Carbon**
	G-14, U.S. Graphite	**MY-10B, Morganite**	**MY-10K, Morganite**	**CY-10C, Morganite**	**CDJ, Morganite**		
Composition	Resin impregnated, phenolic resin	Babbitt impregnated	Antimony impregnated	Resin	Resin, carbonized	Polyester resin	Polyester resin
Apparent density, g/cm^3	1.85	2.6	2.3	1.8	1.75	1.83	1.85
Apparent porosity, %	1.0	1.0	1.0	4.3	7.0	1.0	1.0
Specific heat, kJ/kg · K (Btu/lb · °F)	2093 (0.5)	2846 (0.7)	2093 (0.5)	. . .	. . .	. . .	. . .
Thermal conductivity, W/m · K (Btu/ft · h · °F)	8.0 (4.6)	13.4 (7.8)	13.0 (7.5)	10.5 (6.1)	8.7 (5.0)	8.6 (4.9)	12.0 (7)
Coefficient of thermal expansion, 10^{-6}/K	5.4	5.1	4.7	4.3	2.3	4.9	5.4
Scleroscope hardness	90	75	84	84	95	90	80
Compressive strength, MPa (ksi)	262 (38)	240 (35)	302 (45)	206 (30)	248 (36)	238 (35)	190 (28)
Transverse breaking or flexural strength, MPa (ksi)	77 (11)	88.2 (12.8)	90.2 (13.1)	65.7 (9.5)	59 (8.6)	80 (12)	72 (10)
Tensile strength, MPa (ksi)	58.6 (8.5)	. . .	. . .	. . .	48 (7)	48 (7)	56 (8)
Compressive modulus, GPa (10^6 psi)	. . .	14.51 (2.1)	17.65 (2.6)	11.76 (1.7)	. . .	. . .	. . .
Elastic modulus, GPa (10^6 psi)	19.3 (2.8)	. . .	. . .	. . .	24 (3.5)	14 (2)	23 (3.5)
Flexural modulus, GPa (10^6 psi)	. . .	. . .	. . .	. . .	. . .	. . .	. . .
Maximum service temperature, °C (°F)	260 (500)	120 (248)	350 (660)	300 (570)	. . .	. . .	232 (450)

Seal Friction and Wear Problems

All of the classical friction and wear processes are operative in seals (Ref 11, 12, 13). Ideally, a seal should operate on a hydrodynamic film that is just thick enough to prevent asperity contact and minimize wear. The seal contact pressure should be just enough so that the hydrodynamic film that is formed generates enough pressure to either resist or balance the pressure differential between the process fluid and the atmosphere. With this ideal balance, seal leakage would be minimized, as would friction and wear.

A typical example is provided by a lip seal running on a steel shaft sealing mineral oil, where the lip contact pressure is controlled by a garter spring. As long as the seal is lubricated with oil and the garter spring tension is correct, wear is minimal, occurring only at start-stop conditions, and the seal life is almost infinite. However, when the shaft surface finish becomes too rough, or when abrasives are introduced into the oil, or when the garter spring tension is increased to reduce leakage, the situation changes. Immediately, heat is generated in the fluid under the seal lip, and abrasives carried in the oil or generated as wear debris from the shaft are retained in the seal surface and held against the journal. As more wear debris is generated, the groove being cut in the shaft grows, with an attendant increase in oil leakage.

To either prevent or delay this sequence of events, it is required that the oil be kept clean, the shaft journal have the proper surface finish and hardness, and the garter spring tension be maintained at its proper value. The oil should be filtered so that suspended dirt particles are less than 10 μm (400 μin.) in diameter, the shaft surface finish should be 0.005 μm (20 μ in.) or less, and hardness should exceed a value of 40 HRC.

In this example, major contributions to the friction and wear process were made by heat and abrasion. It is fairly typical of rapid seal wear and failure that several processes frequently occur either simultaneously or sequentially. Because these are also normally dependent, any corrective action must be evaluated for its synergistic effect. In the above example, the generation of only a small amount of heat is sufficient to trigger the failure cycle. The role of abrasion, caused either by surface asperity contact, dirt particles in the oil, or wear debris, was then sufficient to carry the process to completion. Dirt particles ranging from 7 to 30 μm (280 to 1200 μin.) appear to be the most damaging.

Thermoelastic instability is another thermal process that often leads to seal failure (Ref 14, 15, 16). This phenomenon has been observed largely in harder, high-modulus materials that are used in face seals operating at the high contact pressures required to seal large differential pressures at substantial surface velocities. Included among these materials are ceramics such as manufactured carbon graphites, silicon carbide, and tungsten carbide, as well as some cermets. Typically, this phenomenon can occur when carbon graphite is run against high-cobalt materials. During this process, several carbide particles become heated and expand, rising above the surrounding surface, which increases friction, heat generation, and wear. These hot spots progress around the seal face in the carbide-containing material at a rate that is some fraction of the rotational speed, depending on contact pressure, coefficient of friction, and material properties. As a result, heat checking can appear in the harder material.

Heat checking is identified in these materials as fine radial cracks (across the seal face) that are visible to the unaided eye. These cracks are transcrystalline through the hard, brittle carbides. Failure then occurs as particles flake out of the surface or high-velocity water leakage along the cracks erodes the adjacent material.

Abrasive wear is often of concern and results from many causes, including particulates suspended in the lubricating oil, wear debris from inadequate lubrication, products of corrosion, airborne dust, and a rough surface finish. Glass reinforcing fibers used in Teflon seals and asbestos fibers used in braided asbestos packing are capable of cutting the shaft and may produce grooving to significant depths. Other available materials, such as graphite and bronze fibers, have sufficient strength to be used as reinforcements, but will not gall the shaft.

Adhesive wear occurs when the asperities of either two relatively soft materials or a soft against a hard material, often metal, rub against each other with enough heat and pressure to cause welding of the asperity peaks. The peak of the softer (lower shear strength) material then shears off with subsequent motion and forms wear debris.

Examples of materials that like to weld and therefore frequently experience this kind of wear include aluminum against aluminum, titanium against titanium, babbitt against babbitt, copper on copper, and bronze on bronze. Combinations of these materials, such as aluminum or babbitt on bronze, also have this tendency. All of these materials operate successfully against steel shafting when lubricated. Harder material combinations, such as stellite on stellite, are regularly operated against themselves quite successfully.

Cavitation often occurs in lubricating films when a fluid containing entrained air or a gas suddenly passes from a region of high to low pressure, forming bubbles that rapidly expand and then explode (Ref 17). The explosion tears particles of the surface material away and can ultimately remove enough of the material to cause leakage and failure. This process is obviously limited to high pressure and most commonly occurs with oil or water as lubricant.

This form of wear is most common at the inside edge of a high-pressure seal after the rubbing faces become divergent as a result of abrasive or adhesive wear. Once this starts, it is difficult to control. Some seals operate in a regime (Ref 18, 19) in which the fluid changes phase while in the seal face and, hence, may experience similar erosion problems. An example might be a hot water seal in which friction heating of the fluid causes it to flash into steam.

Corrosion is occasionally an operative form of wear in sealing applications. The formation of lead, tin, copper, and iron oxide in seawater can provide the abrasive material to initiate severe abrasive wear. Particular care must be taken in the chemical interactions that can occur between the seal, shaft, housing, and fluid that is present, as well as atmospheric materials, to ensure that corrosive wear is minimized.

Nitrile rubber against mild steel shafting in seawater in the presence of either monel or gunmetal (G-bronze) has been observed to produce severe crevice corrosion under stagnant conditions. Some stagnant adhesion of the rubber to the gland face and shaft has also been observed with subsequent tearing of rubber particles during rotation. Fretting may be present between two surfaces, with one metal having relative vertical/vibratory motion, such as in a valve seat or the joints of a housing around an O-ring. It is usually identified, when leakage increases, by the characteristic color and presence of iron oxide. It can sometimes be eliminated by changing joint materials or inserting a gasket.

Erosion is a frequent, if not preferred, form of wear in seals. It can be self-initiating, as in cavitation erosion, or can take place either along the cracks created by heat checking or in the diverging section of a circumferentially undulating face of a large, high-pressure seal. Particles are removed from the sealing surfaces by the velocity of the stream of fluid and any entrained hard particles. Because the process accelerates with increased gap height and leakage rate, it grows geometrically to an unacceptably high leakage rate and failure. Some materials are more susceptible to erosion than others, depending on their bonding and the grain or particle size.

Design Considerations

The environmental and operational conditions in which a seal must function should determine all of its other characteristics, such as material, geometry, surface finish, and hardness. The best seal for an application is the one that provides effective sealing of the fluid involved at minimum cost and with maximum life.

A number of specific fluid effects must be considered when selecting a seal for a specific application. Some fluids, such as water, are more prone to cavitation erosion than others, whereas fluids such as seawater, strong acids, and bases present more corrosive wear problems. Because water has a low comparative viscosity, it tends to produce a thinner lubricating film at high contact pressures and, hence, generates heat by asperity rubbing, which places high stress on the material.

The chemical compatibility of all of the materials in the shaft-seal-fluid interface is important to seal life. For example, PTFE seals generally perform better in chemical process plants because of their good resistance to attack by most caustic or acidic fluids, whereas natural rubber makes a poor seal in these applications. Special shaft and housing materials, such as stainless steel, monel, Inconel, titanium, and G bronze, must often be used when operating in corrosive environments such as seawater. Although these metals offer better corrosion resistance, they also are subject to crevice corrosion in the same way as elastomeric seals, and corrosion may progress at an accelerated rate when they lose their passivating film.

The operational factors that must be considered in generating a good seal design and in selecting appropriate materials start with the differential pressure that must be sealed. When using mechanical face seals, it is good practice to have the face contact pressure exceed the pressure being sealed by at least 344.6 kPa (50 psi), but to also keep it as low as possible to minimize wear. In a low-pressure lip seal design, a hydrodynamic pressure under the lip that exceeds the differential pressure by only 7 or 14 kPa (1 or 2 psi) may provide effective sealing. In the case of the lip seal, it is also desirable to keep the pressure generated as low as possible in order to minimize leakage. This balance between leakage, contact pressure, wear, and hydrodynamic pressure is delicate and controls the success or failure of a design.

A typical mechanical face seal is shown in Fig. 1. The face contact load is controlled by the sum of the spring load and the hydraulic closing force (Ref 20, 21, 22). The closing force is the hydrostatic pressure multiplied by the active shoulder area of seal ring, plus the total spring force minus the opening force, which is usually taken as the area of the seal ring nose multiplied by one-half the hydrostatic pressure. Shoulders on the outside diameter (OD) of the seal ring are often necessary to accommodate a secondary O-ring static seal. One-half the hydrostatic pressure is used in calculating opening force because it is usual to assume that the pressure decays linearly from the seal ring face OD to the inside diameter (ID). Such an assumption may not be quite correct, because the actual pressure decay is dependent on face geometry (that is, diverging versus converging wedge from OD to ID). This geometry may also change with face wear.

Figure 6 shows the seal ring force balance for an axially floating nonrotating seal ring. The relevant equation for the seal flank closing force is:

$$F_{HC} = P_{HS} \cdot \frac{\pi\,(OD^2 - ID_1^2)}{4} \qquad \textbf{(Eq 1)}$$

where F_{HC} is the hydraulic closing pressure, P_{HS} is the hydrostatic pressure, OD is the seal face outside diameter, and ID_1 is the seal ring skirt inside diameter.

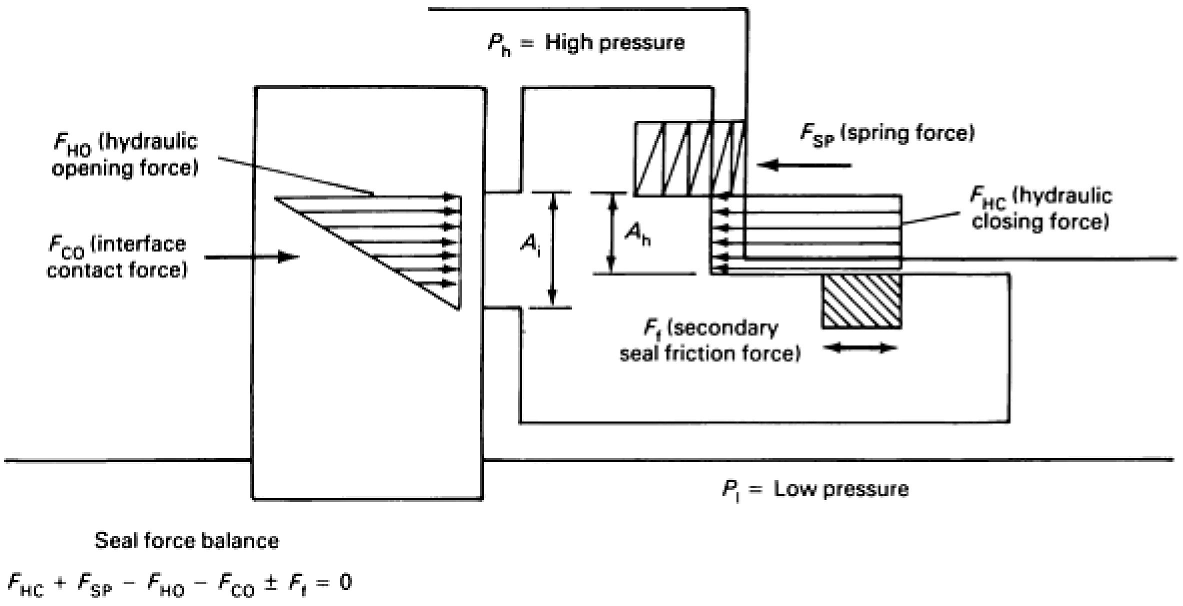


Fig. 6 Face seal force balance. A_i, face contact area, wear area; A_h, unbalanced seal ring area over which hydrostatic pressure acts

For the spring closing force:

$$F_{SP} = N \cdot K_s \cdot \delta \qquad \textbf{(Eq 2)}$$

where F_{SP} is the spring force, N is the number of springs, K_s is the spring rate (lbf/in.), and δ is the spring deflection (in.).

For the face opening force:

$$F_{HO} = \frac{P_{HS}}{2} \cdot \frac{\pi\,(OD^2 - ID_2^2)}{4} \qquad \textbf{(Eq 3)}$$

where F_{HO} is the hydraulic opening force and ID_2 is the seal face inside diameter.

For the face contact pressure:

$$P_c = \frac{4 \cdot (F_{HC} + F_{SP} - F_{HO})}{\pi\,(OD^2 - ID_2^2)} \qquad \textbf{(Eq 4)}$$

where P_c is the seal face contact pressure (psi).

It is obviously better to have the contact pressure increase and decrease directly with the pressure being sealed, and thereby have the lowest possible spring force. This reduces wear and thermal distress at the face (Ref 23). Reducing the hydrostatic closing force and increasing spring force increases wear during those periods when low differential pressure is being contained.

Material Selection

Material selection is usually based on whether the seal is to be static or dynamic, the range of pressures, and the compatibility of the fluid and seal materials. Static joints are usually sealed with a gasket or an O-ring. High-temperature gasket materials include asbestos and soft metals, such as copper, bronze, and aluminum. Steam turbine joints are often sealed using a very smooth, flat joint face bolted tightly together with a bead of Permatex gunned into a groove in the joint. Other high-temperature, formed-in-place joints are made of materials containing sodium silicate or RTV silicone rubber. Erosion and corrosive wear are often failure modes in these joints.

Lower-temperature flanged joints are gasketed with polymeric materials (vinyl, epoxy, and polyester), cork, or even paper sheets. A variety of O-ring materials are also used because of their reliability and availability. Natural and synthetic rubbers are used, although natural rubber usage is decreasing, because of chemical compatibility problems, except in those applications where its low internal damping is needed. All of these materials are capable of sealing both low and high pressures, depending on configuration.

As a dynamic seal, the reliable O-ring is often used to provide both high- and low-pressure sealing on pistons and in such applications as hydraulic cylinder actuator rods. Its widespread usage means that its properties are well-defined and it is readily available from manufacturer catalogs. In high-pressure applications, it is often necessary to use a PTFE back-up ring to prevent extrusion of the O-ring into the clearance space.

There are numerous variations on the O-ring theme for dynamic seals with special characteristics, including quad ring, cap, U-ring, and V-ring seals. The materials of choice include the same polymers used in O-rings, where PTFE, both filled and unfilled, assumes increasing importance. Lip seals in various forms are the seal of choice for low-pressure sealing applications.

The basic materials remain the same, although fillers become more necessary to help hold the lip form. These fillers can be cotton duck, flax, asbestos, glass, metal, and PTFE fibers. Garter and finger springs are also used to hold the lip in contact with the shaft. Fillers that are added to improve the lubricity of the lip include PTFE, graphite, and molydisulfide.

These same materials and combinations are used for packing, which is perhaps the oldest dynamic seal configuration in use. Packing is usually square in cross section and consists of a braided fiber that can be coated for lubricity and strength, and is frequently packed with a lubricant, such as graphite, petrolatum, and/or PTFE.

There are also many specially shaped packings, such as a chevron, which can be used in individual rings or staged as a normal packing. The existence of these packing options is a testimonial to their effectiveness in many applications. In general, packing lacks resilience and therefore loses effectiveness in applications where large radial shaft motions must be tolerated.

This is also true, to a lesser extent, of lip seals, which have difficulty in following radial shaft motions greater than 0.5 mm (20 mils). Both packing and lip seals have been mounted on a bearing-carrier ring, which moves with the shaft and is then sealed by a flexible diaphragm. This arrangement allows the seal to follow the large shaft motion, and has been used successfully in a number of applications.

Manufactured carbon graphites are the materials of choice for gas-turbine gas seals and high-pressure pump seals (Ref 24). Carbon graphite is often run as the soft face against a harder cermet or ceramic material. The carbon graphites are available in a variety of grades with different amounts and sizes of graphite grains dispersed in an amorphous carbon matrix (Ref 25).

Various metallic fillers can also be added to improve counterface polishing. The laminar graphite crystallites provide lubrication as they separate along low-shear-strength cleavage planes. A transfer film of graphite particles coats the counterface, resulting in low friction and wear. After a blank mixture has been converted in an inert gas furnace at 1500 °C (2700 °F), the porosity is sometimes filled with epoxy. The ring may be reheated, resulting in a lower-porosity material, with the filler converted to carbon.

If an even lower porosity is needed, the ring may again be filled with epoxy, but it is usually not reheated to convert the filler to carbon. A carbon-graphite grade, such as G-14 (Fig. 7a), has excellent lubricity and is therefore easy on the counterface surface, because of the large graphite grains. However, it should not be used at surface temperatures above 400 °C (750 °F), because the unconverted organic fillers begin to oxidize above that temperature.

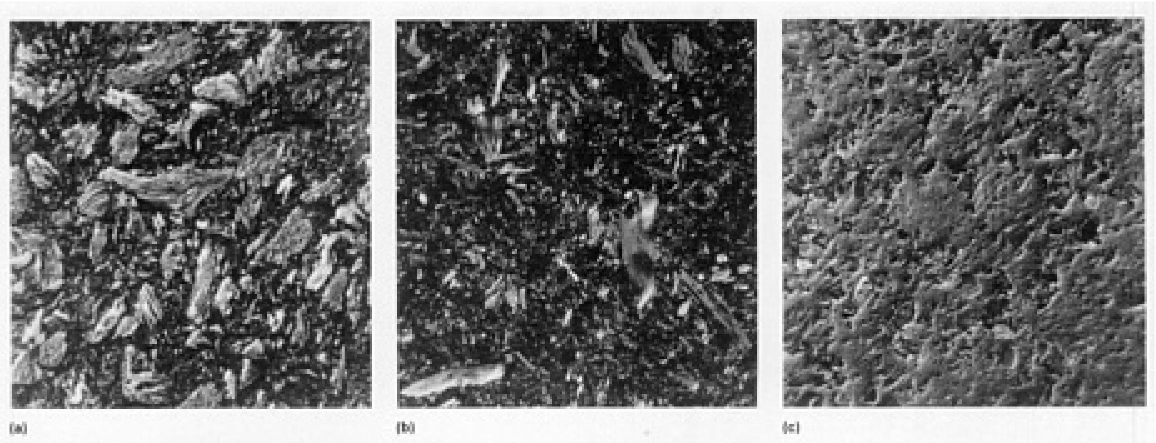


Fig. 7 Photomicrographs of carbon-graphite grades at 175×. (a) Grade G-14. (b) Grade P-658 RC. (c) Grade CDJ

A grade such as P-658 RC (Fig. 7b) has smaller graphite crystals and is stable up to about 550 °C (930 °F), because all the filler has been converted to carbon. It is harder on the counterface in terms of wear. An unfilled carbon, such as CDJ, is stable to 650 °C (1200 °F) and represents the best material for high-temperature operation. Because its graphite grains are small (Fig. 7c), its lubricity is reduced. These examples, photographed in polarized light and magnified 250 times, illustrate the difference in the size of graphite crystallites, although the percentage of graphite in each material is the same.

A carbon material such as G-14 is often run against steel shafting at modest pressures and against ceramic materials at higher pressures, because it is relatively forgiving of the mating material. At higher pressures and surface speeds, heat checking can be a problem for some ceramic material counterfaces, such as stellite. Tungsten carbide performs better as a counterface against unfilled carbons at higher temperatures, but can still encounter heat checking if the contact temperatures become high enough (Ref 25). Silicon carbide appears to offer excellent wear and thermal (heat checking) resistance.

Methods of Reducing Wear

When excessive wear occurs in a specific seal design, it can often be reduced by simply improving lubrication to the contact surface. This can involve either redirecting some of the fluid being sealed or packing the seal lips with grease. In a packing design, a change from one lubricant to another is often enough.

If the seal depends on hydrodynamic lubrication, then the addition of Michells or stepped pockets may significantly reduce wear. If contact pressures are sufficiently high or surface speeds are low enough, it may be necessary to change to a hydrostatic design in order to reduce wear to an acceptable limit (Ref 26).

Coatings have also been effective in reducing shaft wear. Effective coatings that can be applied directly to the shaft or to an intermediate sleeve under the seal include chrome oxide, aluminum oxide, anodized aluminum, nickel, and chrome. A number of very effective self-lubricating materials can be added as fillers to the seal material itself, which not only effectively reduces wear of the seal, but also of the shaft, because they form transfer films on the mating surface. Examples include carbon-graphite, graphite, molydisulfide, and PTFE.

Wear Testing

Three different kinds of wear testing provide useful data about the application of a specific seal design and/or material. Among these are evaluation in the specific operating equipment, bench or rig wear tests, and full-scale simulated tests. Great care is required to ensure that the laws of dynamic similarity are satisfied, that the instrumentation is appropriate and accurate, and that the test machine or equipment operation does in fact cover the range that will be encountered in service. All of these requirements tend to make such tests expensive.

It is customary to plot such test data against *PV*, where *P* is the face contact pressure and *V* is the velocity, and specific or seal nose wear rate for comparative purposes. Specific wear rates for a typical manufactured carbon graphite (see Table 3) rubbing against a silicon carbide hard face (see Table 2) are in the 0.6×10^{-9} to 6×10^{-9} $mm^3/N \cdot m$ range and increase in value with increases in *P* and *V*. If applicable wear data are available in the literature, it is more cost effective to use that data than to build a test machine, because of the problems associated with duplicating long-term equipment operation in a test machine. In such tests, leakage, wear, changes in surface character, temperature, lubricating film thickness, and friction torque can all be measured.

An abundance of sophisticated instrumentation is available to define these characteristics, both statically and dynamically. Metallurgical and scanning electron microscopes provide useful information on the structural and surface material character (Ref 27). Nearly as useful are the electron spectroscopy for chemical analysis (ESCA) and Auger microscopes. Such instruments as the Tallysurf and various laser metrology systems provide convenient means for determining wear and changes in the character of the surface.

Bulk material and fluid temperatures are normally measured by thermocouples or resistance temperature detectors (RTDs), whereas small local contact thermometers, and spectral, laser, and monochronomatic light interferometry are used to determine local surface and hot-spot temperatures. Lubricating film thickness is often measured using proximity gages of the optical, inductive, and capacitive types. All of these provide important information on both seal and material performance. Much of this information is in the open literature, manufacturer publications, and research laboratory reports (Ref 28).

Monitoring

In-service monitoring of seal performance is normally based on leakage. An upper limit is usually set for a specific seal and system, and depends on system size, the fluid being sealed, and equipment disassembly and seal replacement cost. For maintenance decisions, it is often adequate to estimate the quantity of leakage visually. In other cases, the leaked fluid can be collected in a calibrated container and measured volumetrically or it can be weighed to establish leakage per unit time. Often, the pressure differential across the seal is measured using standard pressure transducers. These readings are recorded to provide a time history of any degradation in performance. Less frequently, temperature, vibration, and/or friction torque are measured to assist in the evaluation or elaboration of performance.

References

1. F.J. Dashnaw and L.W. Winn, "Evaluation of Selected Lip Seals/Liners for Propulsion Shafts of Merchant Ships," Paper 22, Propellers '81 Symposium, Society of Naval Architects and Marine Engineers, May 1981
2. H.S. Cheng, C.Y. Chow, and D.F. Wilcock, Behavior of Hydrostatic and Hydrodynamic Noncontacting Face Seals, *J. Lubr. Technol. (Trans. ASME),* Apr 1968, p 510
3. W. Shapiro, "Analysis and Design of Helium-Buffered Face Seals for the SSME High-Pressure Oxygen

Turbopump," Paper 90-2049, American Institute of Aeronautics and Astronautics, July 1990

4. H. Heshmat and W. Shapiro, High-Temperature, Unbalanced, Dry Contact Face Seal, Interfacial Phenomenon and Design Considerations, *Lubr. Eng.,* Apr 1989
5. H.J. Sneck and J.F. McGovern, Analytic Investigation of the Spiral Groove Face Seal, *J. Lubr. Technol. (Trans. ASME),* Apr 1973, p 499
6. W. Shapiro, J. Walowit, and H. Jones, Analysis of Spiral-Groove Face Seals for Liquid Oxygen, *ASLE Trans.,* Vol 27 (No. 3), Apr 1983, p 177-188
7. D.A. Elrod, D.W. Childs, and C.C. Nelson, An Annular Gas Seal Analysis Using Empirical Entrance and Exit Region Friction Factors, *J. Tribol. (Trans. ASME),* Vol 112, Apr 1990, p 196
8. L.P. Ludwig and R.L. Johnson, "Sealing Technology for Aircraft Gas Turbine Engines," Paper 74-1188, American Institute of Aeronautics and Astronautics, Oct 1974
9. W. Shapiro and R. Colsher, Steady-State and Dynamic Analysis of a Jet Engine, Gas Lubricated Shaft Seal, *ASLE Trans.,* Vol 17 (No. 3), Apr 1973, p 190-200
10. H. Heshmat, W. Shapiro, and D.F. Wilcock, Design and Test of a Magnetic-Fluid, Static-Centrifugal Seal, *Lubr. Eng.,* Vol 37 (No. 9), Sept 1981, p 520-526
11. E.H. Iny, A Theory of Sealing with Radial Face Seals, *Wear,* Vol 18, 1971, p 51-69
12. R.B. Snapp and K.R. Sasdelli, "Performance Characteristics of a High-Pressure Face Seal with Radially Converging Interface Shapes," Paper E4, Sixth International Conference on Fluid Sealing (Munich, Germany), Feb 1973
13. E. Mayer, *Mechanical Face Seals,* 5th ed., Butterworths, 1970
14. M. Tseng and R.A. Burton, Thermal Stress in a Two-Dimensional (Plane Stress) Half-Space for a Moving Heat Input, *Wear,* Vol 79, 1982, p 1-9
15. F.E. Kennedy and S.A. Karpe, Thermocracking of a Mechanical Face Seal, *Wear,* Vol 79, 1982, p 21-36
16. F.D. Ju and J.H. Huang, Heat Checking in the Contact Zone of a Bearing Seal (A Two-Dimensional Model of a Single Moving Asperity), *Wear,* Vol 79, 1982, p 107-118
17. R. Haardt, Flow Considerations Around the Cavitation Area in Radial Face Seals, Cavitation and Related Phenomena in Lubrication, *Proceedings of First Leeds-Lyon Symposium on Tribology,* Mechanical Engineering Publications Ltd. (London), Sept 1974, p 221
18. P.A. Beatty and W.F. Hughes, Stratified Two-Phase Flow in Annular Seals, *J. Tribol. (Trans. ASME),* Vol 112, Apr 1990, p 372
19. W. Shapiro and R. Colsher, Selection, Analysis and Preliminary Design of a Steam-Lubricated, Steam-Turbine, Shaft Seal, *ASLE Trans.,* Vol 14 (No. 3), July 1971, p 226-236
20. R.B. Snapp, "Theoretical Analysis of Face Type Seals with Radial Face Profiles," 1964 ASME Winter Annual Meeting (New York), American Society of Mechanical Engineers, Nov 1964
21. W. Shapiro, C. Lee, and H. Jones, "Analysis and Design of a Gas-Lubricated, Sectored, Floating Ring Seal," Paper 87, Trib 55, American Society of Mechanical Engineers, Dec 1987
22. W. Shapiro, "Seal Technology at MTI," Report 90M7, Mechanical Technology Inc., Oct 1990
23. W. Shapiro, "Dynamic Analysis of Contact Face Seals," Paper 860381, Society of Automotive Engineers Conference (Detroit), Feb 1986
24. P.C. Stein, "Circumferential Seals for Use as Oil Seals," ASLE 33 Annual Meeting (Dearborn, MI), American Society of Lubrication Engineers, Apr 1978
25. R.R. Paxton, *Manufactured Carbon: A Self-Lubricating Material for Mechanical Devices,* CRC Press, 1979
26. R.L. Johnson and K. Schoenherr, Seal Wear, *Wear Control Handbook,* American Society of Mechanical Engineers, 1980, p 727
27. A.U. Lebech, "A Mixed Friction Hydrostatic Mechanical Face Seal Model with Thermal Rotation and Wear," Paper 79-AM-41-3, ASLE 34th Annual Meeting (St. Louis, MO), American Society of Lubrication Engineers, Apr 1979
28. F.J. Tribe and G.A. Green, "Assessment of Mechanical Seal Face Materials Under Controlled Interface Torque," ASLE Preprint 85-TC-1C-21, ASLE/ASME Tribology Conference (Atlanta, GA), American

Society of Lubrication Engineers, Oct 1985

Friction and Wear of Internal Combustion Engine Parts

J.R. Davis, Davis & Associates

Introduction

THE INTERNAL COMBUSTION ENGINE is one of the most important types of engineering power plants, encompassing gasoline piston engines used in passenger automobiles, outboard engines for motorboats, and small units for lawn mowers and other such equipment, as well as diesel engines used in trucks, tractors, earth-moving equipment, and locomotives. This article will examine the friction and wear phenomena that deleteriously affect engine performance and fuel economy. Emphasis will be on automotive engines; aircraft engines and reciprocating-type compressor and pump engines are described elsewhere in this Volume. The companion article on "Internal Combustion Engine Lubricants" in this Handbook should also be consulted.

Engine Types and Design Considerations

Principles of Operation (Ref 1, 2). In an internal combustion engine, fuel (gasoline or diesel fluid) is burned within the engine itself (in contrast to a steam engine, where fuel is burned in a separate furnace), and the combustion products serve as the mechanism that converts the latent energy of the fuel into mechanical energy. Characteristic features common to all commercial internal combustion engines include (1) the compression of air, (2) the raising of air temperature by the combustion of fuel in this air at an elevated temperature, (3) the extraction of work (energy) from the heated air and the resultant expansion of pressure, and (4) exhaust.

Figure 1 illustrates these steps schematically. For the intake stroke, the intake valve has opened and the piston is moving downward, drawing air and gasoline vapor into the cylinder. During the compression stroke, the intake valve has closed and the piston is moving upward, compressing the mixture. During the power stroke, the ignition system produces a spark that ignites the mixture. As it burns, high pressure is created, which pushes the piston downward. For the exhaust stroke, the exhaust valve has opened and the piston is moving upward, forcing the burned gases from the cylinder. As will be described, interaction of the moving parts during the engine cycle results in a complex tribosystem.

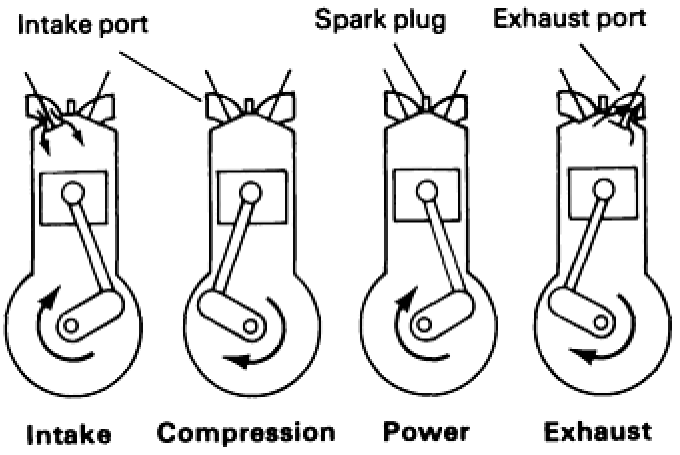


Fig. 1 Schematic of a four-stroke engine cycle. Source: Ref 1

Engine Types (Ref 2, 3). There are four types of internal combustion engines: four-stroke cycle engines (Fig. 1), two-stroke engines, rotary (Wankel) engines, and continuous-combustion gas turbine engines. Because they are the most common, four- and two-stroke engines will be described in this article. Both of these engine designs, which are based on the reciprocating piston principle, include spark-ignition and compression-ignition types. In spark-ignition (Otto cycle) engines, the fuel and air are mixed in either the carburetor or intake port (fuel injection) prior to entry into the combustion chamber. Gasoline is the most commonly used fuel in spark-ignition engines, although commercial gases such as blast-furnace gas, coal gas, coke-oven gas, and natural gas are used in large stationary engines. In compression-ignition engines, the intake system supplies only air, which is compressed before the fuel is injected into the combustion chamber. Compression-ignition engines use liquid of low volatility, such as fuel oil or diesel oil.

Four-stroke compression-ignition and spark-ignition engines are used in most automotive applications. Depending on the cylinder arrangement, four-stroke engines can be subcategorized as in-line (Fig. 2), V (vee), opposed piston, and radial types. Two-stroke spark-ignition engines are commonly used in lightweight applications, such as outboard motors, motorcycles, and snowmobiles, whereas two-stroke compression-ignition diesel engines are used in on-high-way trucks, buses, low-speed marine applications, and some railroad locomotives.

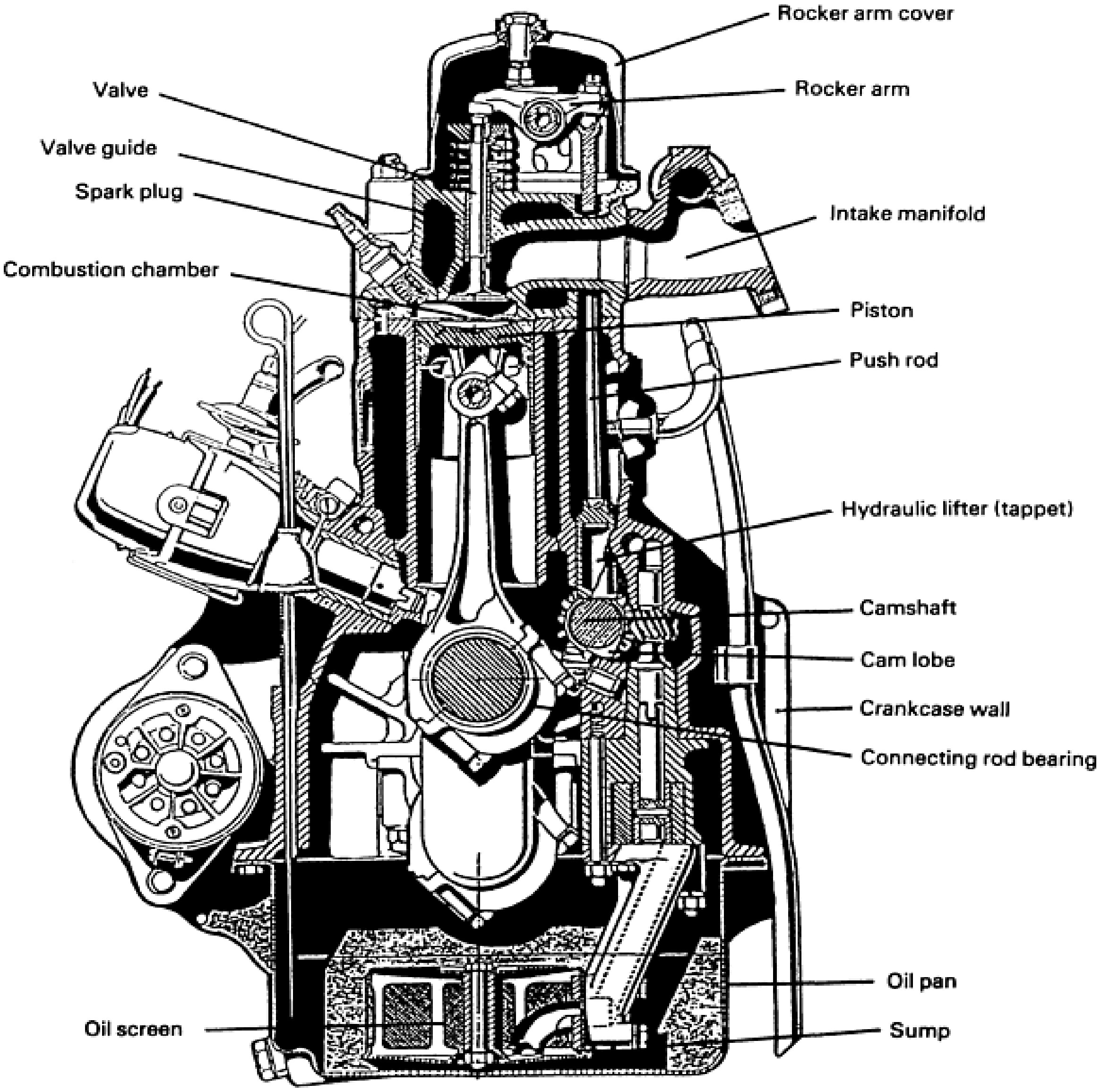


Fig. 2 Cross section of a six-cylinder, in-line automotive engine. Courtesy of S.Q.A. Rizvi, Lubrizol Corporation

General Features. Automobile engines have been produced with various numbers of cylinders over the years (from 2 to 16), but the most common engines in modern cars are in-line four-cylinder designs for smaller cars and in-line six-cylinder, V-6, and V-8 designs for larger cars. Smaller engines, both in terms of cylinder number and size (displacement), provide improved fuel economy. The displacement of modern passenger cars ranges, on average, from 1.6 to 5.7 L (98 to 350 in.^{3}).

All engines employ compact, rigid structures. Most cylinder blocks are made from gray cast iron, although aluminum is increasingly common, particularly in cylinder heads. Four-cylinder engines use three or five crankshaft bearings; the V-6 uses four, the V-8 uses five.

Liquid-cooling passages are provided around the bore of each cylinder as well as in the hot regions of the combustion chambers and exhaust ports in the cylinder head. Modern pistons are generally made from aluminum alloys (see the section "Pistons and Piston Ring Assembly" in this article), but gray iron or malleable iron pistons are used in heavy-duty diesel applications. Crankshafts are usually made from cast iron (gray iron or ductile iron) or forged steel. Similarly, connecting rods are made from cast iron (malleable iron), conventional steel forgings, or powder metallurgy steel forgings.

Detachable cylinder heads contain the inlet and exhaust valves (overhead valves). Some engines employ a single exhaust and inlet valve, while other designs feature three or four valves per cylinder. Exhaust gas is directed through short passages to the exhaust collector or manifold.

Fuel and air are delivered to the cylinder head by a compact inlet manifold, usually made from aluminum. Fuel management is provided by either a carburetor or an electronic fuel injection system. In most cases, feedback from an exhaust gas combustion sensor is used to provide precise control of the fuel/air ratio and to optimize the performance of the emission control system. Electronic ignition virtually eliminates the need for periodic ignition system maintenance.

Many of the engine components mentioned above are shown in Fig. 2. Detailed information on the friction and wear problems associated with some of these components is provided below.

Designers of advanced fuel-efficient, low-emissions engines are beginning to include ceramic parts because of weight and durability considerations. While the relatively high cost of ceramics may limit their current use, improvements in material cost, processing efficiency, and machining may eventually result in the use of increasing numbers of ceramic parts in the internal combustion engine.

Friction and Wear of Engine Components

Since the oil and gas shortages in the 1970s, significant automotive engineering efforts have been devoted to emission and fuel economy improvements in order to meet legislated emissions and Corporate Average Fuel Economy (CAFE) standards. These efforts include (Ref 4):

- Vehicle downsizing for reduced weight and rolling resistance, and exterior design improvements for reduced aerodynamic drag
- Engine downsizing to reduce the power-to-weight ratio (see earlier comments regarding engine displacement)
- Engine design improvements in weight reduction, volumetric efficiency, and improved combustion
- Drive-train improvements, for example, lockup torque converter, split-torque converter, overdrive, and reduced parasitic friction losses of drive-train components (see the article "Friction and Wear of Automotive and Truck Drive Trains" in this Volume)
- Advancements in exhaust treatment to meet emission standards, for example, catalyst technology
- Fuel metering with electronic, closed-loop engine control to maximize fuel economy and minimize emissions at various engine operating conditions
- Reductions in chassis friction and tire rolling resistance (see the article "Friction and Wear of Tires" in this Volume)
- Use of lightweight, durable ceramics and ceramic coatings

Fuel economy goals have also resulted in renewed interest in the understanding of engine friction and wear in the design of engine components with improved tribological performance. For example, recent studies have indicated that a 6.9 kPa (1 psi) mean effective pressure (MEP) reduction in engine friction yields a 1 to 2% fuel economy improvement in four-cylinder automobiles (Ref 4).

Contribution of Major Components to Engine Friction. At a typical engine speed of the Environmental Protection Agency (EPA) metro-highway (M-H) driving cycle, the major engine friction components in order of importance are (Ref 4):

- Piston ring assembly
- Valve train
- Crankshaft (bearings)
- Oil pump

Figure 3 shows the percentage of friction for these various components as a function of speed. The percentage of friction increases with increasing speed for the crankshaft bearings, connecting rod bearings, and pumps; the friction percentage decreases for the piston and ring assembly and the valve train assembly (Ref 5).

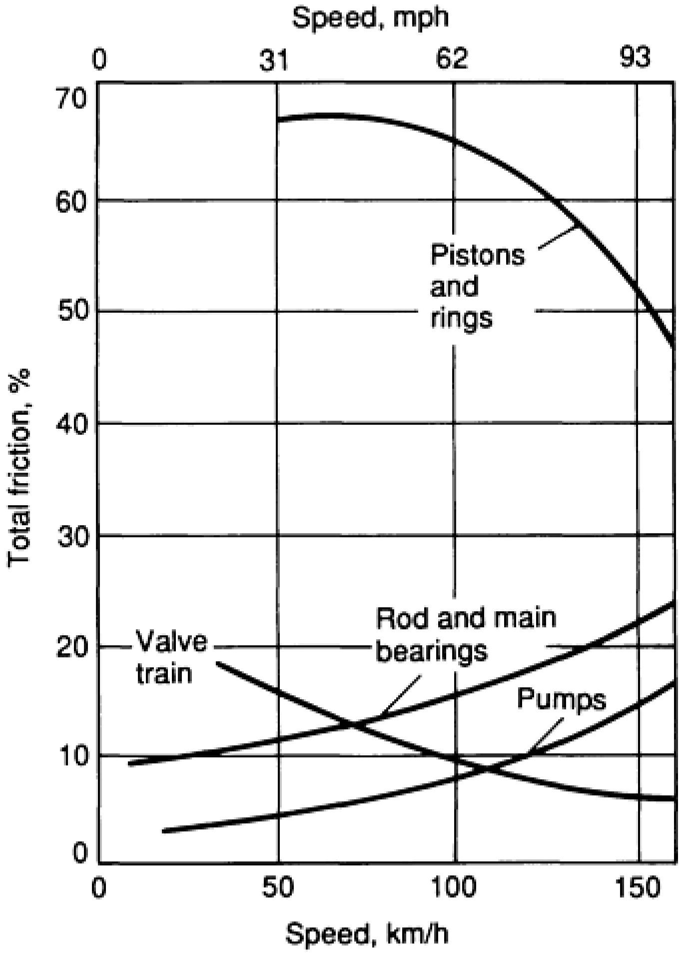


Fig. 3 Effect of speed on engine component friction. Source: Ref 5

Lubrication Regimes for Friction Components (Ref 6). A difficult problem in understanding engine friction is the wide variation in the magnitude of the friction forces observed--the coefficient of friction ranges from 0.2 for boundary lubricated engine components to 0.001 for hydrodynamically lubricated engine components. Figure 4 relates the various regimes of lubrication to the major engine friction components. The critical elements in the valve train, including

the cam-lifter interface and rocker-arm pivots, operate in the boundary to mixed regime, because high loads are encountered at low sliding speeds. Piston rings operate over the range of thick-film hydrodynamic to mixed lubrication, because the gas load and velocity of the piston ring vary extensively over one engine cycle. At top dead center of the piston stroke when the gas forces are maximum, the velocity of the piston ring is zero and contact occurs between the piston ring and the cylinder wall. The piston skirt, because of its large contact area and low loading, is primarily a thin-film hydrodynamic contact. Engine bearings operate in the hydrodynamic mode, except for brief periods during starting and stopping. More detailed information on the fundamentals of engine lubrication can be found in the article "Internal Combustion Engine Lubricants" in this Volume.

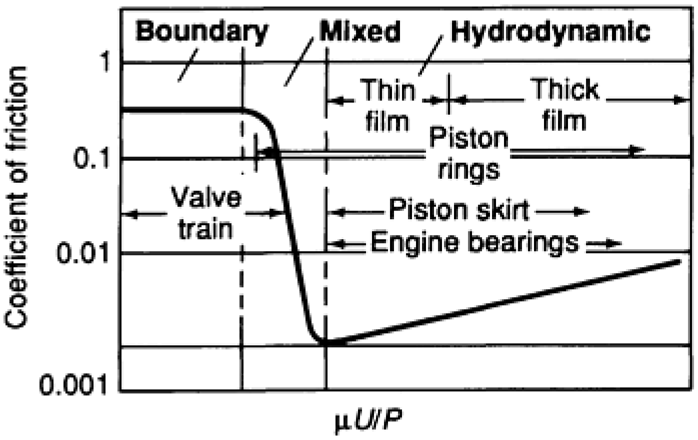


Fig. 4 Stribeck diagram showing lubrication regimes for major engine friction components. μ**, absolute viscosity;** ***U*****, speed;** ***P*****, unit load. Source: Ref 6**

Engine Wear (Ref 7). Normally, all reciprocating engines operate with lubrication. Oil of the wrong viscosity or oil that deteriorates in service can have a decided influence on wear of cylinders, cylinder liners, rings, camshafts, and valve trains.

The products of combustion of the fuel contain water vapor that carries potentially corrosive gases, such as oxides of sulfur, carbon, and nitrogen. If allowed to condense on engine parts, these oxides readily form acids and cause excessive corrosive wear. As an example, the relation between percentage of sulfur in the fuel and the wear of piston rings is shown in Fig. 5(a) and (b). Cylinder wear is similarly affected.

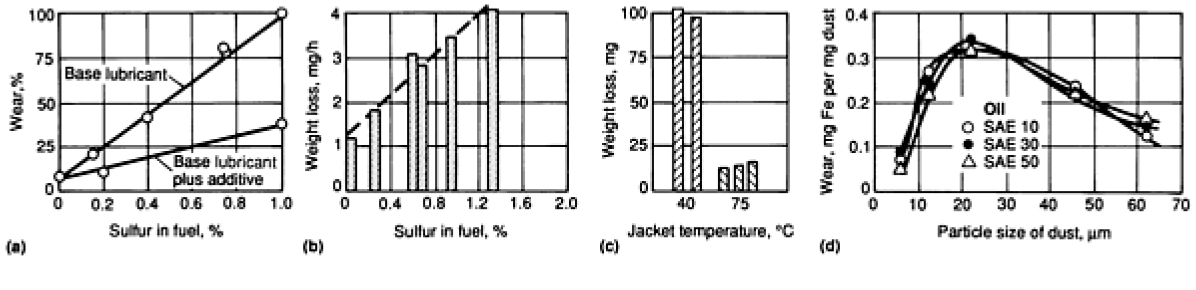


Fig. 5 Effect of variables on the wear of piston rings. Source: Ref 7

Corrosive wear is minimized by running the engine hot, by hastening warmup, and, especially with diesel engines, by using low-sulfur fuel. The results shown in Fig. 5(c) are typical in that they show reduced piston ring wear at temperatures above 65 °C (150 °F); wear of cylinders is similar. Also, lubricant additives will often inhibit corrosive effects, as illustrated in Fig. 5(a). The effects of lubricant additives are also described in the article "Lubricant Additives and Their Functions" in this Volume.

Other mechanisms associated with engine wear include:

- Adhesive wear (cylinder liners and piston rings)
- Abrasive wear (cams, rocker arms, and tappets)
- Scuffing (cylinder liners and piston rings)
- Pitting caused by fatigue wear (cams and tappets)

Pistons and Piston Ring Assembly

Among the many moving parts in an internal combustion engine, the piston and piston ring assembly are considered to be the components that contribute most to total engine friction and wear. It is estimated that piston ring friction amounts to approximately 50 to 70% of the friction loss of the piston assembly (Ref 8), and that the piston assembly is responsible for about 25% to as much as 75% of total engine friction losses (Ref 9, 10). References 11 and 12 are useful reviews of theoretical and experimental investigations related to piston ring tribology in the areas of friction and wear determination, friction reduction, lubrication, and oil consumption. Some of the findings reported in those papers are outlined below.

Pistons (Ref 13)

As stated earlier, most pistons for automotive applications are made from aluminum alloys. The universal acceptance of aluminum alloy pistons by gasoline engine manufacturers can be attributed to their light weight and high thermal conductivity.

Aluminum automotive pistons generally are permanent mold castings. The alloy most commonly used for passenger car pistons is 332.0-T5, an aluminum-silicon alloy that has a good combination of foundry, mechanical, and physical characteristics, including low thermal expansion.

Piston alloys for heavy-duty engines include 336.0-T551, a low-expansion alloy, and 242-T571 because of its higher thermal conductivity and superior properties at elevated temperatures. Chemical compositions and typical properties of cast aluminum piston alloys are given in Tables 1 and 2, respectively. A cross section of a piston is shown in Fig. 6.

Table 1 Chemical compositions of cast aluminum piston alloys

Alloy	Composition, wt%									
	Si	**Fe**	**Cu**	**Mn**	**Mg**	**Cr**	**Ni**	**Zn**	**Ti**	**Al**
336.0	11.0-13.0	1.2	0.50-1.5	0.35	0.7-1.3	. . .	2.0-3.0	0.35	0.25	bal
242.0	0.7	1.0	3.5-4.5	0.35	1.2-1.8	0.25	1.7-2.3	0.35	0.25	bal

Table 2 Typical properties of cast aluminum piston alloys

Alloy	Temper and product form(a)	Specific gravity(b)	Density(b)		Approximate melting range		Electrical conductivity, %IACS	Thermal conductivity at 25 °C (77 °F), W/m · K	Coefficient of thermal expansion, per °C × 10^{-6} (per °F × 10^{-6})	
			kg/m^3	**lb/in.^{3}**	**°C**	**°F**			**20-100 °C (68-212 °F)**	**20-300 °C (68-570 °F)**
336.0	T551(P)	2.72	2713	0.098	540-570	1000-1060	29	0.28	18.9 (10.5)	20.9 (11.6)
242.0	T571(P)	2.81	2823	0.102	525-635	980-1180	34	0.32	22.5 (12.5)	24.5 (13.6)
332.0	T5(P)	2.76	2768	0.100	520-	970-1080	26	0.25	20.7 (11.5)	22.3 (12.4)

(a) P, permanent mold.

(b) The specific gravity and weight data in this table assume solid (void-free) metal. Because some porosity cannot be avoided in commercial castings, their specific gravity or weight is slightly less than the

theoretical value.

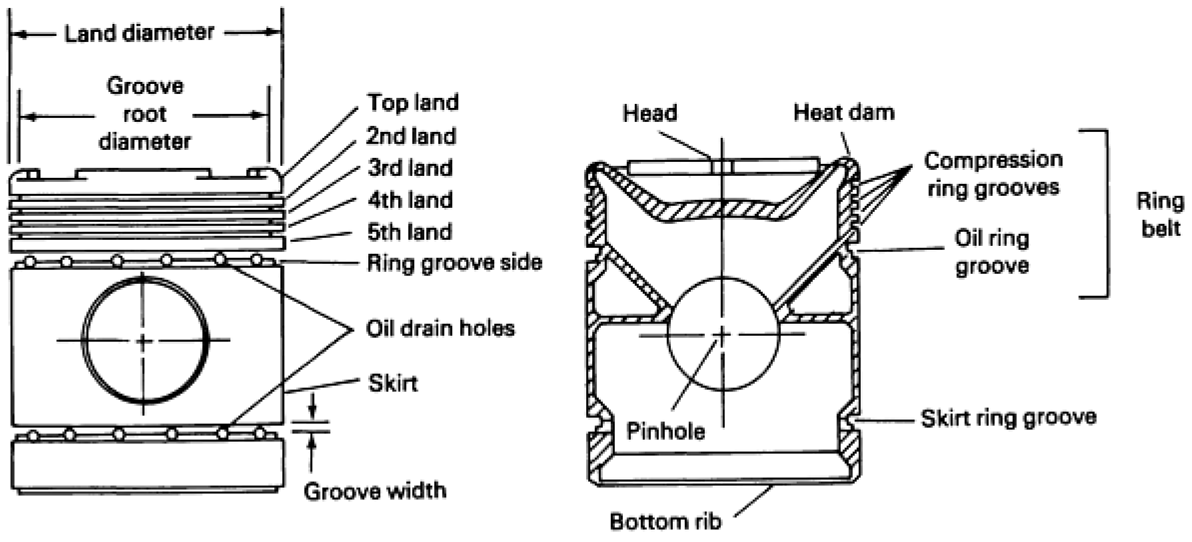


Fig. 6 Typical piston configuration per ANSI/ASME Standard B205

Piston Rings (Ref 11)

Piston rings are mechanical devices used for sealing pistons, piston plungers, reciprocating rods, and so forth. In gasoline and diesel engines, they are generally split-type, self-expanding metal rings. When they are placed in the grooves of the piston (Fig. 6) and provided with a lubricant, a moving seal is formed between the piston and the cylinder bore (liner).

Piston rings are divided into two categories: compression rings and oil-control rings. Compression rings, generally two or more, are located near the top of the piston (Fig. 7). Compression rings have to perform two basic functions--gas sealing and oil control--under the most hostile conditions, being exposed to high temperatures, high gas pressures, extreme stresses, impact, corrosion, and abrasion. They must be able to operate with a minimum of lubrication and still provide service at low wear conditions. Therefore, the basic design considerations are efficient sealing, light weight, and good material strength and minimum wear under elevated-temperature conditions.

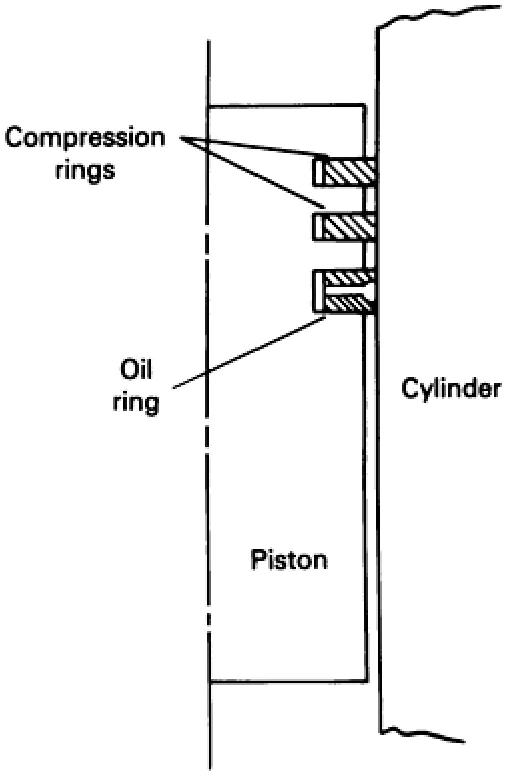


Fig. 7 Schematic of piston ring assembly showing location of compression rings and oil ring

Oil rings, generally one or more, are placed below the compression rings to prevent the passage of excessive lubricating oil into the combustion chamber, yet provide adequate lubrication for the compression rings. Major factors affecting this basic function are ring-bore contact pressure, ring-bore conformability, sliding surface characteristics, and drainage for the surplus oil. Other factors that influence oil ring design include amount of oil transported, oil viscosity, and engine operating and temperature conditions.

Ring Materials and Coatings (Ref 11). An ideal piston ring, material must meet the following requirements: low friction and wear losses, superior scuffing resistance, tolerances for marginal lubrication and rapidly varying environments, good running-in wear behavior, long-term reliability and consistency of performance, long maintenance-free life, and low production costs. Although no single material can meet all of these requirements, base materials exhibiting good strength, fatigue, impact, and wear properties can be coated with materials that have excellent wear, friction, and scuff-resistance characteristics.

Gray cast iron, with a hardness ranging from 200 to 400 HB, is the most commonly used material for compression and oil rings. Table 3 shows the various types of cast irons used in gasoline engine piston rings. Pearlitic gray iron produced by either centrifugal or sand casting methods has the widest use. In heavy-duty engine applications, alloy cast iron, ductile iron, and malleable iron are used. Harder materials, such as carbon steel or even bearing steels, may also be used.

Table 3 Typical piston ring materials

Type	Approximate specification	Manufacturing method	Range of composition, wt% (balance iron)									Tensile strength		Modulus of elasticity		Hardness, HB
			C	**Si**	**Mn**	**S**	**P**	**Ni**	**Cr**	**Mo**	**Other**	**MPa**	**ksi**	**GPa**	**psi × 10^6**	
Gray cast iron	BSS 4K6	Centrifugally cast	3.5 max	1.8-2.5	1.2 max	1.0 max	0.40-0.65	0.4 max	0.5 max	0.4 max		245	36	117	17	210 min
	BSS 4K6	Sand cast	3.5 max	1.0-1.8	0.6-1.2	0.15 max	0.25-0.60	0.4 max	0.4 max	0.4 max		245	36	103	15	200-245
	DTD 233A	Centrifugally cast	3.1-3.4	2.1-2.5	0.7-1.0	0.08 max	0.40-0.65	. . .	0.3-0.6	0.5-1.0		310	45	110	16	255-296
Carbidic/malleable iron	DTD 485A	Individually cast	2.7-3.3	2.0-3.0	0.5-0.9	0.1 max	0.5 max	. . .	0.5-0.85	. . .	V, 0.4 max	585	85	159	23	270-320
	DTD 485A	Centrifugally cast	2.75-3.3	1.8-2.5	0.6-1.0	0.12 max	0.3-0.5	0.3 max	0.65-1.15	0.7-1.0	Al, 0.5 max	400	58	155	22	269-302
Malleable/nodular iron	. . .	Centrifugally cast	3.0-3.3	1.0-1.4	0.6-0.9	0.1 max	0.1 max	0.3 max	0.1-0.5	. . .		540	78	155	22	200-400
	BS2789 SNG	Sand or centrifugally cast	Varies according to mechanical properties required									570	83	165	24	240-297

Source: Ref 11

Relatively thick coatings (up to 0.2 mm, or 0.008 in.) of plated chromium on the ring periphery provide the best compromise between scuffing, wear, and corrosion resistance and low friction and oxidation resistance at high temperatures. Generally, the use of chromium-plated top rings (with a hardness of 700 to 900 HV) run against cast iron cylinder liners can reduce the ring and liner wear by a factor of 2 to 3. Rings coated with flame- or plasma-sprayed coatings of molybdenum (in thicknesses up to 0.25 mm, or 0.01 in.) with a hardness of greater than 1000 HV are believed to have higher scuffing resistance than chromium-plated rings. The major limitation of molybdenum ring coatings is that they are subject to oxidation at 500 °C (930 °F), and at 730 °C (1345 °F) the oxide volatilizes.

Ceramic (Cr_2O_3), cermet (WC-Co, Cr_3C_2-Ni, and Cr_3C_2-Mo), and alloy (Mo-Cr-Ni) plasma-sprayed coatings have also been developed for achieving improved wear and scuffing resistance under conditions of marginal lubrication. Reference 14 reviews various ring materials and coatings.

Cylinder Liner Materials (Ref 11)

The cylinder walls are stressed mechanically by the high gas pressure and the side thrust of the piston, as well as thermally by the high gas temperature. Because all these stress-induced factors are cyclic, the cylinder liner materials must have good mechanical and fatigue strengths; otherwise, cylinder bore distortion or early material fatigue failure may result. In addition, tribological properties such as wear and scuff resistance must be satisfactory, because metal-to-metal contacts between piston rings and the cylinder liner do occur.

Most cylinder liners are made of gray cast iron. To increase their mechanical strength, additions of nickel, chromium, copper, and molybdenum are required. The addition of at least 0.3 wt% Cr to improve wear resistance is fairly common and is considered to be particularly beneficial when combined with 0.65 wt% Mo. Typical liner materials are shown in Table 4.

Table 4 Typical cylinder bore materials

<table>
<tr><th rowspan="2">Type</th><th colspan="8">Composition, wt% (balance iron)</th><th colspan="2">Ultimate tensile strength</th><th rowspan="2">Hardness, HB</th><th rowspan="2">Microstructure</th></tr>
<tr><th>C</th><th>Si</th><th>S</th><th>P</th><th>Mn</th><th>Ni</th><th>Cr</th><th>Other</th><th>MPa</th><th>ksi</th></tr>
<tr><td>Sand-cast blocks and barrels</td><td>3.3</td><td>2.1</td><td>0.1</td><td>0.15</td><td>0.6</td><td>0.3</td><td>0.2</td><td>. . .</td><td>220</td><td>32</td><td>200</td><td rowspan="2">Flake graphite, pearlite matrix, no free carbides. Phosphide eutectic network increases with phosphorus content. Minimum of free ferrite to minimize scuffing, but less important with increasing phosphide</td></tr>
<tr><td>Sand-cast liners</td><td>3.3</td><td>1.8</td><td>0.1</td><td>0.25</td><td>0.8</td><td>. . .</td><td>0.4</td><td>. . .</td><td>230</td><td>33</td><td>220</td></tr>
<tr><td>Centrifugal cast gray iron liners</td><td>3.4</td><td>2.3</td><td>0.06</td><td>0.5</td><td>0.8</td><td>. . .</td><td>0.4</td><td>. . .</td><td>260</td><td>38</td><td>250</td><td>As for sand cast, but with finer graphite tending toward rosette or undercooled. Matrix martensitic/bainitic if liner is hardened and tempered</td></tr>
<tr><td>Centrifugal cast alloy iron liners</td><td>3.1</td><td>2.3</td><td>0.06</td><td>0.3</td><td>0.8</td><td>. . .</td><td>0.8</td><td>0.3 V or 1.0 Mo</td><td>320</td><td>46</td><td>280</td><td>Compact graphite, pearlite matrix with islands of alloy carbides. Ternary eutectic phosphide with carbides. Minimum of free ferrite ideal but not important in presence of carbides</td></tr>
<tr><td>Austenitic iron liners</td><td>2.9</td><td>2.0</td><td>0.6</td><td>0.3</td><td>0.8</td><td>14.0</td><td>2.0</td><td>7.0 Cu</td><td>190</td><td>28</td><td>180</td><td>Fine flake graphite with some undercooled graphite. Fine-grained cored austenite matrix. Complex carbides and ternary phosphide eutectic in broken network</td></tr>
</table>

Source: Ref 11

Steel cylinder liners also have been used, offering the advantage of much thinner walls. They must be hardened to at least 400 HB, however, for satisfactory resistance to wear and scuffing, or else be chromium plated. It should be noted that chromium-plated liners should not be run against chromium-plated piston rings, but rather against plain cast iron rings or rings with other types of coatings.

Aluminum alloy A390.0 cylinder liners with high silicon content (~17% Si) have also been used. Some engine test results show that the liner wear of cylinders made from this material is lower than that of conventional cast iron liners (Ref 11, 14).

The surface finish of the cylinder bore is another key factor influencing oil consumption, scuff and wear resistance, and engine life (Ref 15, 16, 17, 18). Generally a clean, burr-free, consistent surface finish is required. An average surface finish of 0.5 to 0.8 μm (20 to 30 μin.) is recommended. More detailed information on bore finish specifications can be found in Ref 16.

Piston Ring-Cylinder Liner Scuffing (Ref 11)

For piston rings and cylinder liners, scuffing is primarily the result of the breakdown of the lubricating oil film separating the sliding surfaces. This leads to metal-to-metal contact, causing local welding between the ring and the liner.

Piston ring scuffing is generally a running-in problem and is particularly likely to occur if full load and speed are applied too rapidly. It can also occur after longer running periods if the lubrication conditions deteriorate. The symptoms of scuffing in an engine are usually high oil consumption and blow-by, and sometimes a noticeable engine knocking noise. If scuffing is allowed to proceed, it may cause piston seizure or a major increase in piston clearance, resulting in a heavy knocking noise. However, in some cases the prime scuffed area does not spread, but heals itself, and the engine is still able to perform satisfactorily. Some of the methods used to reduce the incidence of scuffing include:

- Fitting the piston rings with the running surface preshaped to a profile resembling that obtained after running-in so that less running-in wear is required
- Providing a surface finish on the cylinder liner that is coarse enough to increase the wear rate of the rings. If the rings are preshaped, however, the finish should be made finer than if they are flat
- Using a small amount of mildly abrasive material as a coating or filling on the liner or rings that will help to maintain an adequate initial wear rate
- Increasing the oil supply to the upper rings during running-in, which helps flush out wear debris and counteract any local deficiencies in the oil distribution

Factors Affecting Piston Ring Wear (Ref 7)

The principal factors that influence wear of piston rings are speed, temperature, load, frequency of use, dirt, corrosion, surface finish (see above), and quantity of lubricant.

Speed influences ring wear largely as a function of piston travel per kilometer. The rate of ring wear in a high-speed engine of a given power rating is higher than that of a low-speed engine of the same rating in proportion to the distance traveled by the rings. Design weaknesses of rings, pistons, and cylinders become more evident as speed increases.

Although radial face wear is the prime consideration, wear of ring and groove sides cannot be ignored. Engine speed affects side wear because it affects the acceleration forces on the rings.

Temperature. Although high temperature occasionally causes wear by interfering with lubrication of cylinder surfaces, low temperature is a far more frequent source of excessive piston and ring wear. As coolant temperature falls below 65 °C (150 °F), cylinder and ring wear increases rapidly because of corrosion caused by condensate that carries corrosive combustion products, as shown in Fig. 5(c). When it is impossible to maintain proper temperature, lubricating oils that contain additives are helpful (see the article "Lubricant Additives and Their Functions" in this Volume).

Engine load affects the wear rates of both cylinders and rings principally through its influence on temperature and corrosive wear. Engines in light-load service generally have a high rate of corrosive wear. High loads seldom cause wear problems unless design weaknesses cause scuffing because of distortion or destruction of the oil film by hot spots.

Frequency of Use. Engines operated the most miles per year have the lowest rates of cylinder and ring wear. Low annual mileage usually means short runs, more starts per mile, and detrimentally low operating temperatures; the sensitivity of an engine to this type of wear is largely determined by its warmup rate.

Airborne dirt is a chronic cause of excessive ring wear. Proper design and maintenance of filter equipment is essential, as is prevention of air leaks between filters and the engine. Dirt in the lubricating oil is a less frequent problem but, again, proper design and maintenance of oil filters and seals are vital.

Abrasive Particle Size. When the particle size of dirt in an engine reaches a certain size, wear subsides and the dirt particles often reduce friction rather than increase it. Test data in Fig. 5(d) for piston ring wear indicate that abrasion is most severe at a particle size of about 20 μm.

Reducing Piston Ring Assembly Friction

Piston assembly design modifications have resulted in substantially reduced engine friction. As shown in Fig. 4, the light loads and large contact areas associated with the piston skirt promote a hydrodynamic contact. In order to reduce a hydrodynamic friction, skirt areas have been reduced in recent designs. For example, the slipper-type piston skirt design is typically preferred to the full trunk design. Skirt length can have a significant impact on friction level (Ref 5, 19), as shown in Fig. 8. Decreasing the effective skirt length from the original design length reduces friction and improves engine economy for both low-speed, low-load and high-speed, higher-load operation.

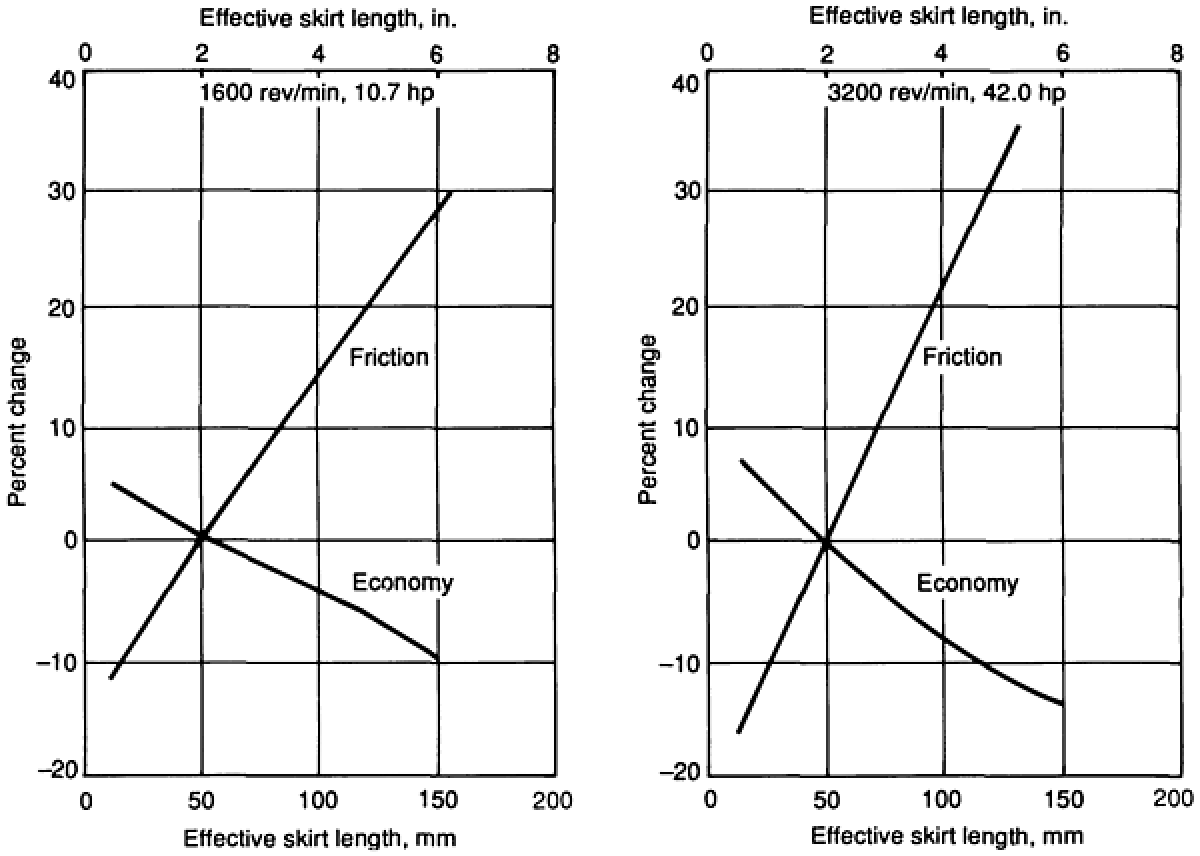


Fig. 8 Effect of piston skirt length on friction and fuel economy. Source: Ref 6

Recent developments in finite-element modeling of pistons have led to a better understanding of operating skirt profiles (Ref 20). These developments, coupled with piston dynamic analysis, can be used to optimize piston skirt design for reduced friction.

Ring size and ring tension effects on friction have been examined in a movable-bore engine (Ref 21). This engine was a single-cylinder diesel engine containing four piston rings. Data from these tests show that substantial reductions in ring friction can be realized by reducing ring tensions and face widths. References 4, 6, and 12 provide detailed information on tests that have been developed for determining general friction considerations for piston assembly design.

Valve Train Assembly

The valve train assembly includes such components as the camshaft, cams, tappets (hydraulic lifters), valves, valve seat rings, valve springs, and rocker arms. As the camshaft rotates, the cams use pushrods and rockers to operate the cylinder valves, which admit the fuel or expel the exhaust gases. In an overhead camshaft engine, the cams lie above the valves and move them directly. It is estimated that as much as 25% of total engine friction losses are attributable to the valve train assembly (Ref 6). Contributions of specific components to valve train friction torque loss are shown in Fig. 9 (Ref 22).

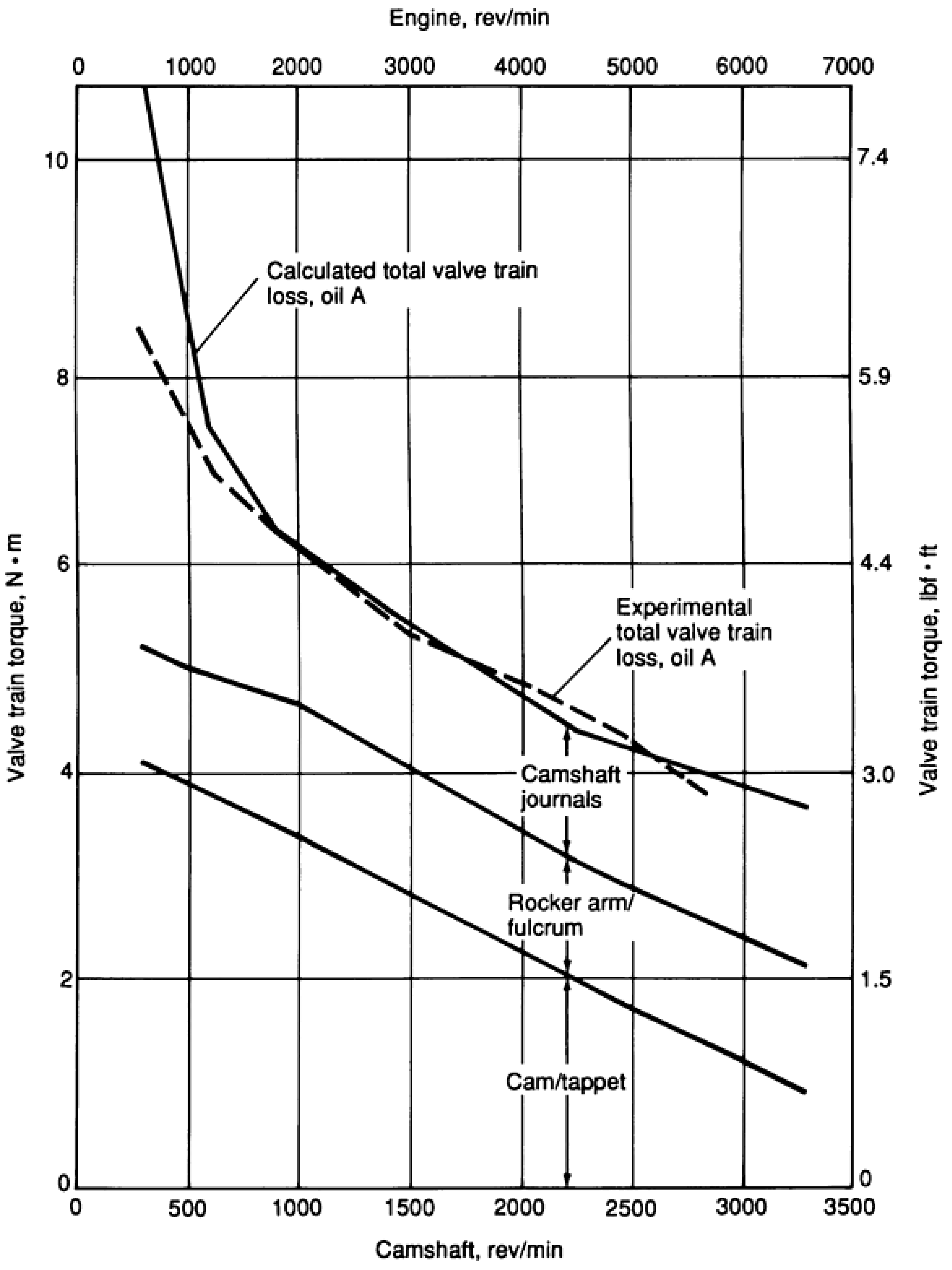


Fig. 9 Total torque loss and torque loss for the critical valve train friction contacts versus speed calculated for a

non-friction-modified SAE 30 oil at 100 °C (212 °F). Source: Ref 22

Valve Train Friction (Ref 6). Valve train design presents a challenge because high loads must be carried over the entire speed range of the engine. Loads acting on the valve train at low speeds are caused primarily by valve spring forces, while at higher speeds inertia forces generated by component mass dominate. Several valve train designs are shown in Fig. 10, including the pushrod, the direct-acting overhead cam, the end-pivot overhead cam, and the center-pivot overhead cam (Ref 23). Boundary lubrication is the predominant lubrication regime in the valve train (Ref 24), as shown in Fig. 4, and the major energy losses occur in the cam-lifter interface and in the rocker-arm pivot. As shown in Fig. 11, the torque required to drive the valve train is lowest for the direct-acting design. This is because there is only one sliding contact. The other design alternatives exhibit high friction because of the additional rocker-arm pivot loss.

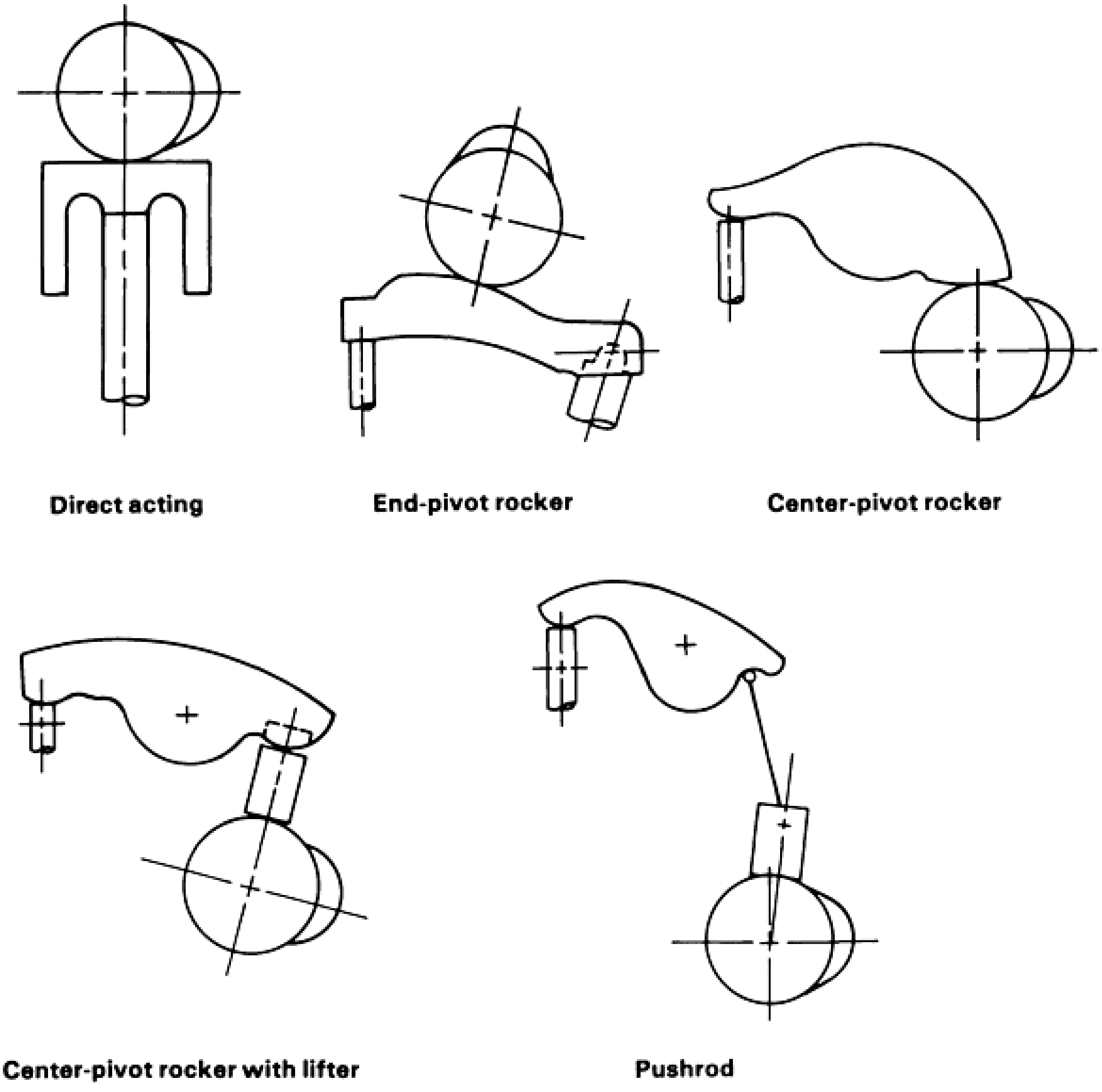


Fig. 10 Valve train designs. Source: Ref 23

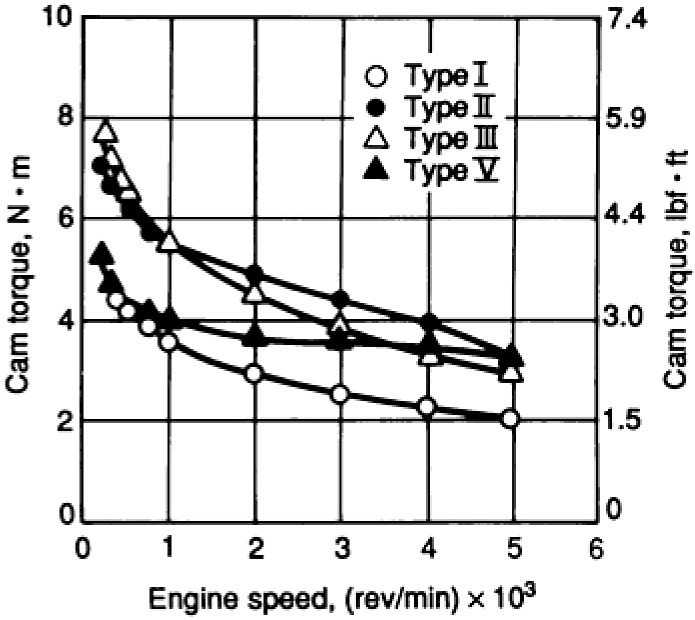


Fig. 11 Energy losses in various valve train designs. See also Fig. 10. Source: Ref 6

One effective method of reducing valve train friction is to use rolling-element bearings at critical locations. Such bearings are well suited to carry high loads at low speeds. Such designs (Fig. 12) have been used in racing engines. Rolling-element bearings are included at the cam-lifter interface, the rocker-arm pivot, and the interface between the rocker nose and the valve stem. Studies have shown that the valve train torque can be reduced by as much as 50% at 1500 rev/min using rolling-element bearings; the projected vehicle fuel improvement is 2.9% (Ref 4). Cam roller-followers of silicon nitride and partially stabilized zirconia are beginning to find application in diesel and automotive engines because they can replace a needle-bearing-type roller with a single, simpler rolling element, thereby reducing the weight of the engine. Disadvantages of the rolling-element bearing design include the susceptibility of the bearings to rolling-contact fatigue (see the articles "Friction and Wear of Rolling-Element Bearings" and "Rolling Contact Wear" in this Volume) and increased complexity and cost.

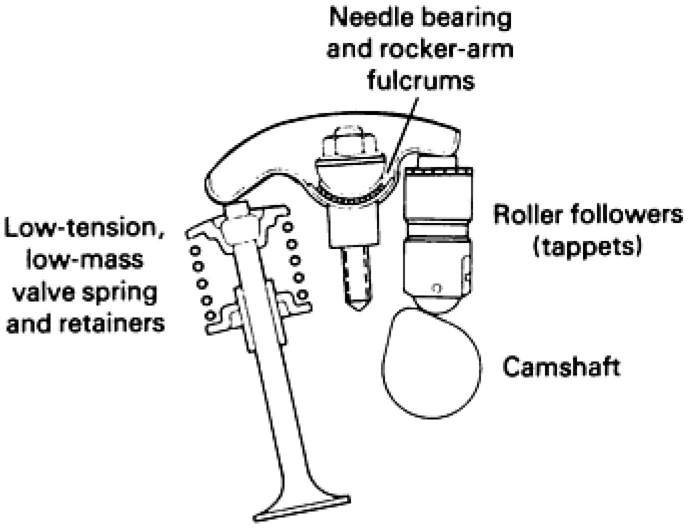


Fig. 12 Low-friction valve train design incorporating rolling elements. Source: Ref 22

An alternative to the rolling-element bearing approach is a reduction in valve spring load (Fig. 13). The projected improvement in fuel economy is about 0.5% (Ref 4).

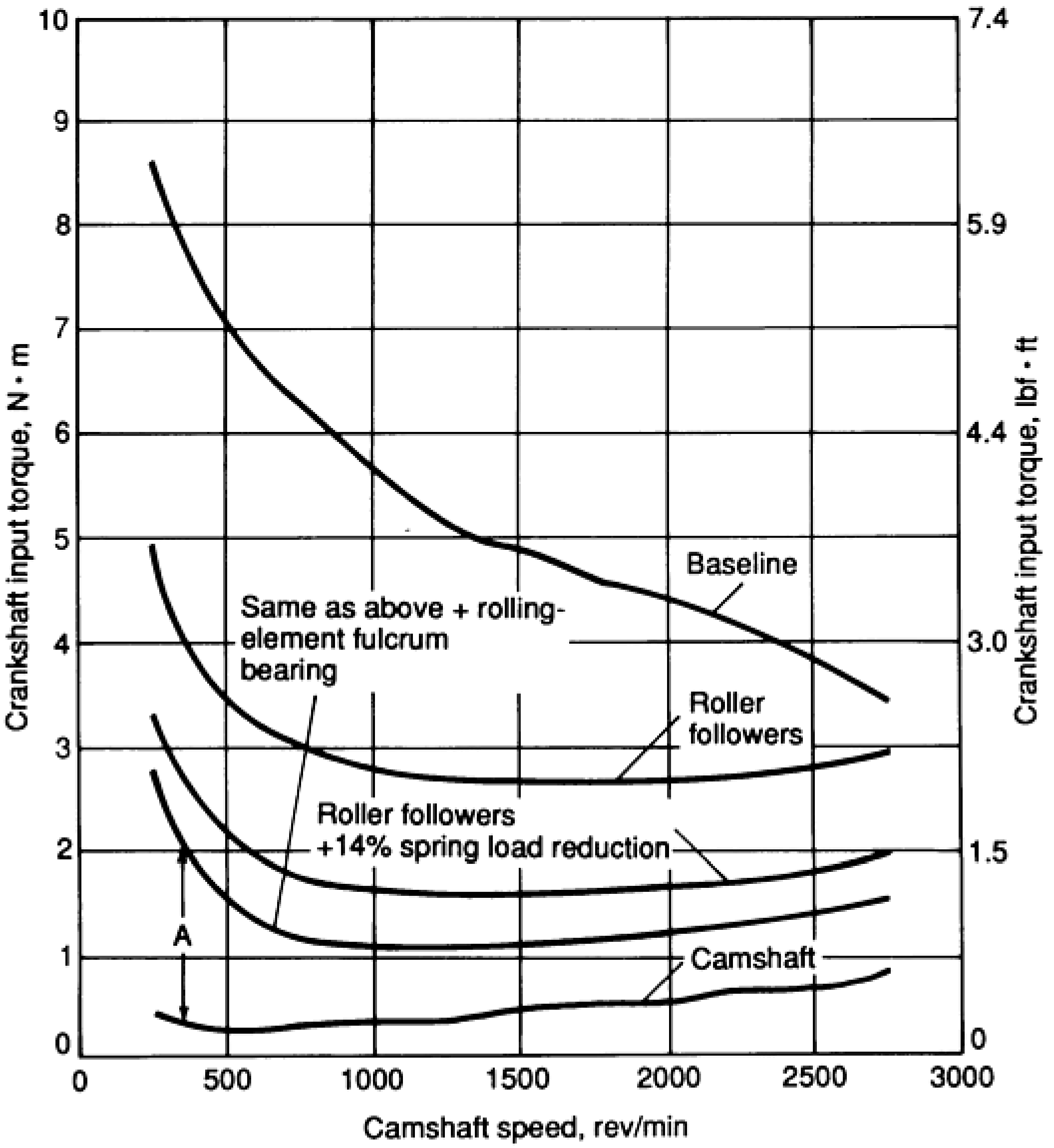


Fig. 13 Effect of reduced valve spring load on camshaft friction. Included in region A are valve guides, seals, rocker arm tips, fulcrum, tappets, and rollers. Source: Ref 6

A number of studies have also been conducted on valve train friction in terms of lubrication principles (Ref 22, 25, 26). These analyses have shown that oil viscosity has a relatively small effect on overall friction loss, but that a marked reduction in friction can be obtained with oils containing friction modifiers, or additives. Results from these studies (see Table 5 and Fig. 14) suggest that selection of the proper lubricant may reduce friction levels to those obtainable by component redesign.

Table 5 Lubricant characteristics for valve train friction analysis

See also Fig. 14.

Oil	Viscosity, cSt		Elemental composition, ppm					Measured friction coefficient
	40 °C (105 °F)	**100 °C (212 °F)**	**Zn**	**P**	**Mo**	**Mg**	**B**	
A	76.3	9.2	1308	1185	. . .	1190	141	0.1
B	269.5	20.5	1767	1612	. . .	1496	169	0.1
C	77.6	9.3	1243	1099	1998	998	125	0.05

Source: Ref 22

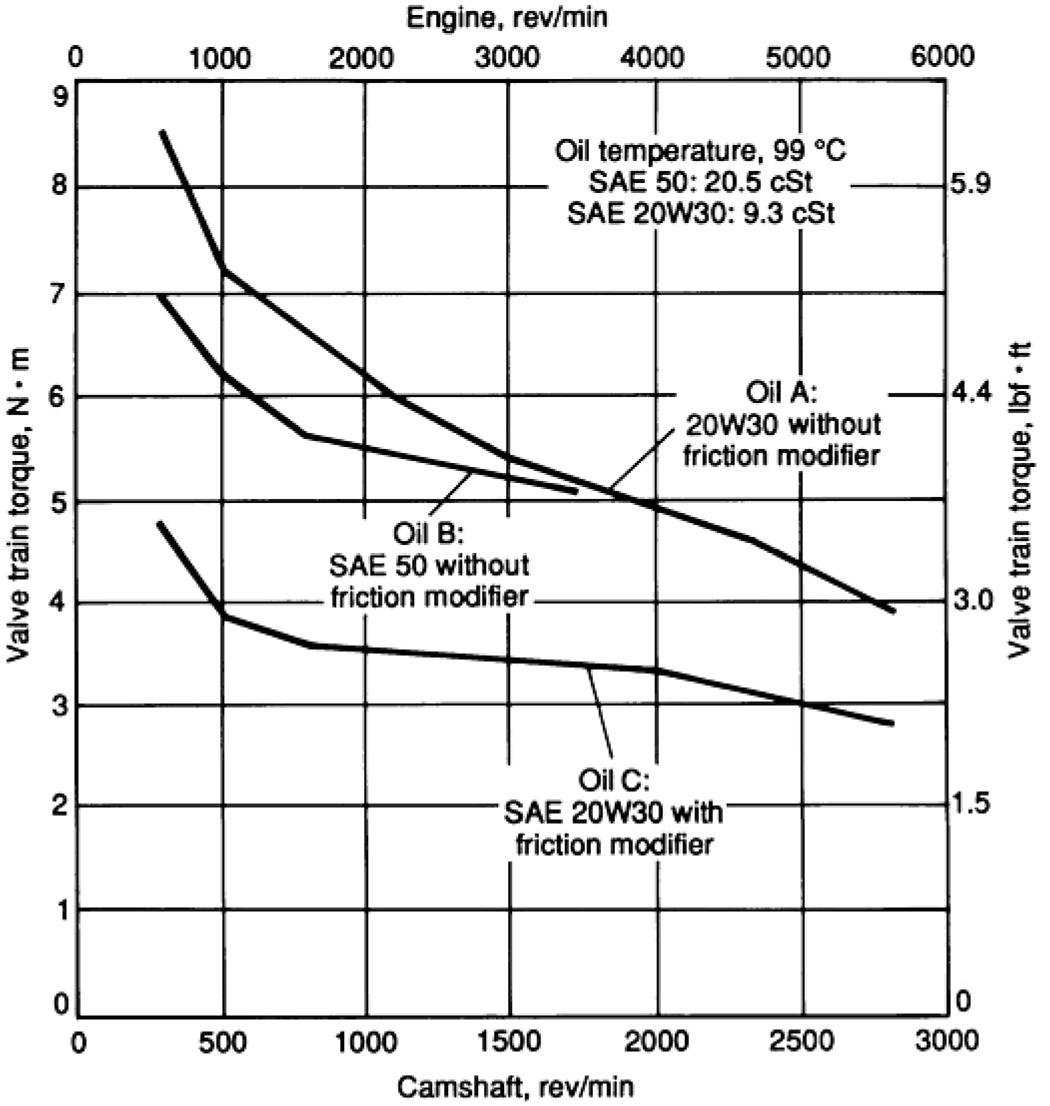


Fig. 14 Effects of oil viscosity and of a friction modifier on friction torque loss. See also Table 5. Source: Ref 22

Valve Train Wear. As with friction, the cam-tappet interface is also the most critical wear consideration. Under heavy duty, cams and tappets suffer from burnishing (due to adhesive/abrasive wear processes), scuffing (due to severe adhesive wear processes), and pitting (due to fatigue wear processes) (Ref 27). The wear modes for cams and tappets are very similar to those for gears (see the article "Friction and Wear of Gears" in this Volume).

The wear of cams and tappets can be reduced considerably by selecting hard material combinations or by hardening the cam material by heat treatments, thermochemical treatments, or by applying coatings. Hardenable gray iron camshafts (SAE Grade G4000d) are the most widely used; steels may be water-quenched carbon steels, oil-quenched alloy steels of

0.5 to 0.7% C, or a carburizing grade. The most common tappet material for automotive applications is gray cast iron of composition similar to the following: 3.20% C, 2.25% Si, 0.80% Mn, 1.10% Cr, 0.60% Mo, 0.55% Ni, 0.20% max P, and 0.10% max S (Ref 7). Ceramic materials such as silicon nitride are finding increased use as tappet shims and wear pads.

Types of coatings and surface treatments that can be adapted for cam and tappet materials to reduce wear include (Ref 27):

- *Running-in coatings,* such as phosphate coatings, chemically produced oxide coatings on ferrous metals, and electrochemically deposited tin and aluminum
- *Hard Surfaces.* Surfaces of cams can be hardened by diffusion treatment, such as carburizing, nitriding, and Tufftriding
- *Hard coatings,* such as TiN and TiC, can be applied by physical vapor deposition or chemical vapor deposition to cams and tappets to achieve low coefficients of friction and wear

Silicon nitride valves, while currently expensive, have been used very successfully in high-performance racing engines. Their low mass and high stiffness allow them to be used at the high speeds (>10,000 rev/min) typical of such engines.

Crankshaft Bearings (Ref 6)

Large loads are carried by hydrodynamic journal bearings (Fig. 4) with low energy losses during the engine cycle because of the complete separation of the two rubbing surfaces by a fluid film. Analysis of crankshaft bearing friction has been carried out by computer solutions, which calculate complete bearing performance (Ref 28, 29). Loads on crankshaft journal bearings vary in magnitude and direction because they are composed of inertial loads caused by compression and expansion events.

Bearing clearance and journal diameter are two important parameters that are controlled during the engine design process. The effect of bearing clearance on calculated bearing performance is shown in Fig. 15. As clearance increases, the bearing friction loss decreases, while the minimum film thickness exhibits a local maximum. Durability considerations suggest that to prevent bearing wear, maximum film thickness must be maintained. Thus, most journal bearings are designed to provide film thicknesses on the order of 2.0 μm. The friction power that could be saved by increasing journal bearing clearance beyond this maximum film thickness point is overshadowed by requirements for maintaining an adequate minimum oil film thickness.

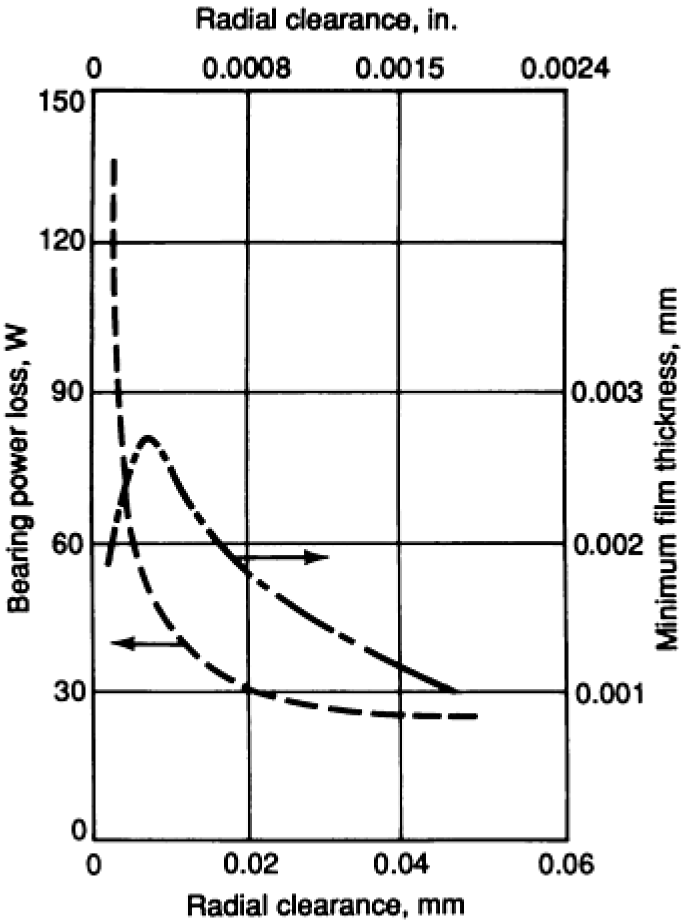


Fig. 15 Effect of bearing clearance on bearing performance. Source: Ref 6

Journal bearing diameters also significantly affect bearing power loss. Studies have shown that decreasing the rod journal diameter results in a 3 to 4% reduction in engine friction and a corresponding 1% improvement in fuel economy (Ref 19). It should be pointed out, however, that bearing diameter also controls the torsional strength of the crankshaft. Therefore, a design decision must be made between decreasing bearing diameters to reduce friction and the resulting loss in crankshaft strength.

Oil Pumps (Ref 4)

The types of oil pumps generally used in engines are the spur gear, internal-external gear, and tip-sealing gear (Fig. 16). These gears are typically driven at engine speed or half engine speed. Pump choice and drive ratio are usually dictated by available design space (especially length) and the selected drive mechanism. Once the location and drive are determined, noise, pressure fluctuations, cost, and efficiency trade-offs are considered.

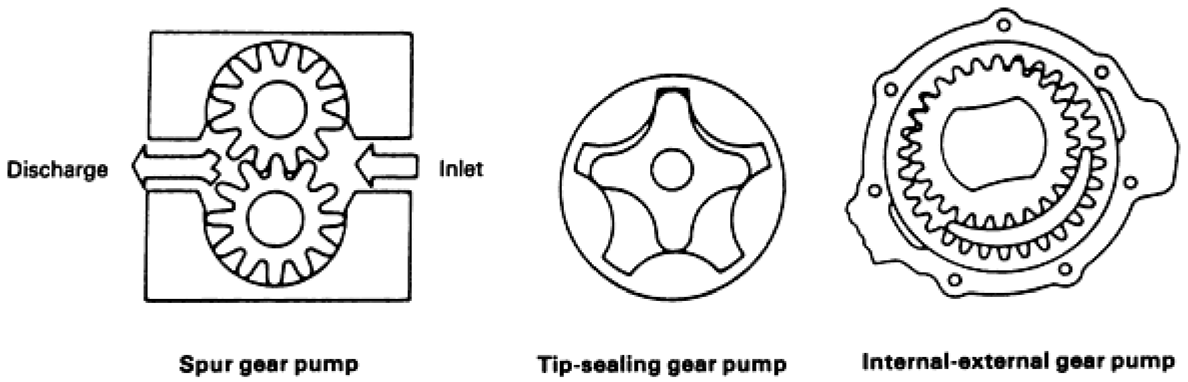


Fig. 16 Oil pump gear designs. Source: Ref 4

Engines with the camshaft in the block typically have a common drive shared by the distributor and oil pump; spur gear of tip-sealing pumps are used in such applications. These pump designs have relatively small diameters, which minimize friction losses between the gear outer diameter and the housing. Engines with overhead camshafts have no convenient half-speed oil-pump drive, especially if the distributor is driven from the camshaft. In this case, the designer must choose between providing an auxiliary shaft at extra cost or installing the pump on the crankshaft. The crankshaft option is usually selected for cost considerations, but the length of the pump is normally limited. This in turn forces selection of a thin pump (usually an external-internal gear type) with a large-diameter gear to provide the required displacement. The diameter of the outer gear is a function of the inner gear size required to go over the crankshaft. The large diameters of crankshaft-driven pumps result in higher losses because of the increased frictional torque. Figure 17 compares the input requirements measured for different pump designs. Fuel economy gains of 0.6% can be attained through proper pump design.

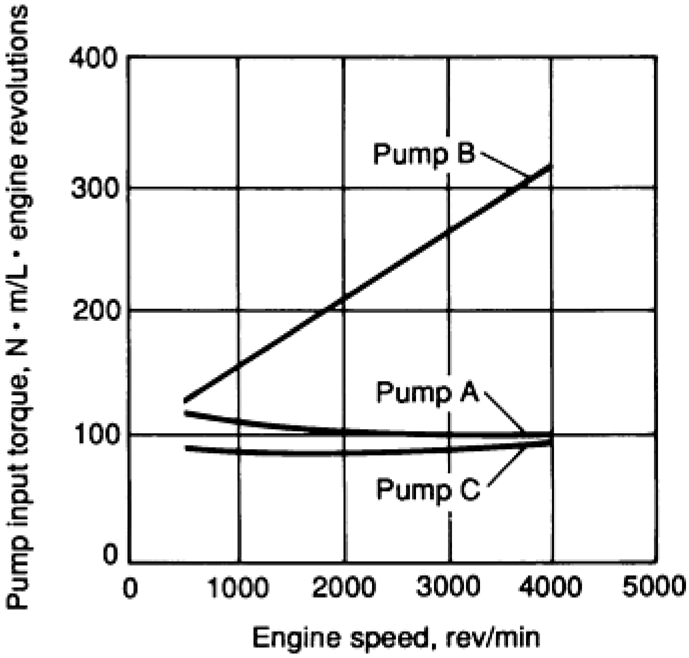


Fig. 17 Normalized oil pump power consumption. Pump A: tip-sealing gear, half-speed drive, 67 mm (2.6 in.) diameter. Pump B: internal-external gear, engine speed drive, 100 mm (4 in.) diameter. Pump C: tip-sealing gear, half-speed drive, 41 mm (1.6 in.) diameter. Source: Ref 4

Future Outlook

As new alternative fuels and more efficient engine designs are introduced, ceramics and ceramic coatings and composites will replace some of the traditional engine materials. Part of the reason for this substitution is that alcohol-based fuels are naturally less lubricious than traditional fuels and more durable surfaces are required to support their use. Aside from the cam roller-followers and valves mentioned above, other candidate parts for the use of ceramics include valve guides, tappet shims, turbocharger rotors, fuel injector pins, head plates with integral valve seats, and exhaust port liners. Despite the increased use of advanced ceramic materials, steels, cast irons, and aluminum alloys and their composites will remain important structural materials for the internal combustion engine.

References

1. S.P. Parker, Ed., *McGraw-Hill Encyclopedia of Science & Technology,* McGraw-Hill, 1982, p 545
2. E.A. Avallone and T. Baumeister III, Ed., *Marks' Standard Handbook for Mechanical Engineers,* 9th ed., McGraw-Hill, 1987, p 9-80 to 9-85
3. E.F. Obert, *Internal Combustion Engines and Air Pollution,* Intext Educational Publishing, 1968, p 1-25
4. J.T. Kovach, E.A. Tsakiris, and L.T. Wong, "Engine Friction Reduction for Improved Fuel Economy," Paper 820085, Society of Automotive Engineers, 1982
5. A.E. Cleveland and I.N. Bishop, "Several Possible Paths to Improved Part-Load Economy of Spark-Ignition Engines," Paper 150A, Society of Automotive Engineers, 1960
6. R.C. Rosenberg, "General Friction Considerations for Engine Design," Paper 821576, Society of Automotive Engineers, 1982
7. T. Lyman, Ed., *Metals Handbook,* Vol 1, 8th ed., American Society for Metals, 1961, p 250, 251
8. L.C. Lichty, *Internal Combustion Engines,* McGraw-Hill, 1951
9. O. Pinkus and D.F. Wilcock, *Strategy for Energy Conservation Through Tribology,* American Society of Mechanical Engineers, 1971
10. J.A. McGeehan, "A Lubrication Review of the Effects of Piston and Ring Friction and Lubricating Oil Viscocity on Fuel Economy," Paper 780673, Society of Automotive Engineers, 1978
11. L.L. Ting, Lubricated Piston Rings and Cylinder Bore Wear, *Wear Control Handbook,* American Society of Mechanical Engineers, 1980, p 609-665
12. L.L. Ting, "A Review of Present Information on Piston Ring Tribology," Paper 852355, Society of Automotive Engineers, 1985
13. A. Kearney and E. Rooy, Aluminum Foundry Products, *Metals Handbook,* Vol 1 2, 10th ed., ASM International, 1990, p 123-151
14. B.J. Taylor and T.S. Eyre, A Review of Piston Ring and Cylinder Liner Materials, *Tribology Int.,* April 1979, p 79
15. R. Munro and G.H. Hughes, "Piston Ring Application in Diesel Engines," Paper 8, presented at Associated Engineering--Technical Symposium, June 1980
16. D.M. Hesling, "A Study of Bore Finishes and Their Effects on Engine Performance," presented at American Society of Mechanical Engineers Annual Meeting, April/May 1963
17. J.E. Willn, Characterisation of Cylinder Bore Surface Finish--A Review of Profile Analysis, *Wear,* Vol 19, 1972, p 143
18. J.C. Campbell, Cylinder Bore Surface Roughness in Internal Combustion Engines: Its Application and Control, *Wear,* Vol 19, 1972, p 163
19. I.N. Bishop, "Effects of Design Variables on Friction and Economy," Paper 812A, Society of Automotive Engineers, 1964
20. P. Reipert and R. Buchta, "New Design Methods for Pistons," Paper 810933, Society of Automotive Engineers, 1981
21. S. Furuhama, M. Takiguchi, and K. Tomizawa, "Effect of Piston and Piston Ring Designs on the Piston Friction Forces in Diesel Engines," Paper 810977, Society of Automotive Engineers, 1981
22. J.T. Staron and P.A. Willermet, "An Analysis of Valve Train Friction in Terms of Lubrication Principles,"

Paper 830165, Society of Automotive Engineers, 1983

23. W.B. Armstrong and B.A. Buuck, "Effect of Various Valve Gear Designs on Valve Gear Energy Consumption," Paper 810787, Society of Automotive Engineers, 1981
24. A. Dyson, Kinematics and Wear Patterns of Cam and Finger Follower Automotive Valve Gear, *Tribology,* June 1980, p 121-132
25. P.A. Willermet, J.M. Pieprzak, and D.P. Dailey, "Friction Reduction in Valve Trains: The Influence of Friction Reducing Oil Additives," Paper 890725, Society of Automotive Engineers, 1989
26. P.A. Willermet and J.M. Pieprzak, "Some Effects of Lubricant Composition and Tappet Rotation on Cam/Tappet Friction," Paper 89-Trib-2, presented at ASME/STLE Joint Tribology Conference (Fort Lauderdale, FL), 16-19 Oct 1989
27. B. Bhushan and B.K. Gupta, *Handbook of Tribology,* McGraw-Hill, 1991, p 16.10-16.13
28. A.O. DeHart and D.H. Harwick, "Engine Bearing Design: 1969," Paper 690008, Society of Automotive Engineers, 1969
29. P.E. Vickery, "Friction Losses in Automotive Plain Bearings," Paper 750052, Society of Automotive Engineers, 1975

Friction and Wear of Automotive and Truck Drive Trains

William Kelly, Borg-Warner Automotive, Inc.

Introduction

NORMAL WEAR CHARACTERISTICS of several major automotive power-train components are described in this article. The wear descriptions provided here are intended as sources of useful reference information for:

- Evaluating components in vehicle performance analyses
- Identifying minimum acceptance criteria in component endurance and durability testing
- Delineating a general indication of the expected operating life for major power-train elements in normal service applications

It should be emphasized that the issues addressed below represent those wear characteristics found in normal operating use. Typical operating life estimates have therefore been included. Abuse, neglect, and external influences beyond those normally associated with typical land vehicle operation are too varied in their effects on wear and too complicated to be considered within the scope of this article.

Empirical data, when available, are presented. Information deemed proprietary by the source, relative to specific wear-resistance methods, has been protected by normalized graphical methods, where required. However, the results of those wear-resistance preventative measures are detailed as much as possible. Typical wear-resistant features and designs are described for elements such as thrust surfaces and high-torque transmission components. The wear discussions in this article are confined to macroscopic views of wear effects. Applications are categorized by main vehicle component areas, and include automatic and manual transmissions, transfer cases, engine components, and wheel end components.

Automatic Transmission Components

The major source of wear for automatic transmission components has always been attributed to energy dissipation in the friction elements. The typical placement and use of friction products in an automatic transmission are shown in Fig. 1.

Clutch bands and clutch plates make up the friction element interface of the major wear components. The methods employed to minimize the wear effects on these friction surfaces are described in the section "Clutch Bands and Plates" in this article. Wear effects that result from seal degradation and the subsequent abrasive contamination are observed as a constant source of concern. However, from the normal wear perspective, seal wear associated with a properly designed and applied sealing system should not be significant during the normal operating life. A review of the normal maintenance procedures for automatic transmissions, where seal replacement is not specifically identified or required, supports the preceding statement. In addition, the proper maintenance and replacement of the automatic transmission fluid (ATF) or gear lubricant should not result in appreciable wear contributions, provided that adequate maintenance intervals are identified.

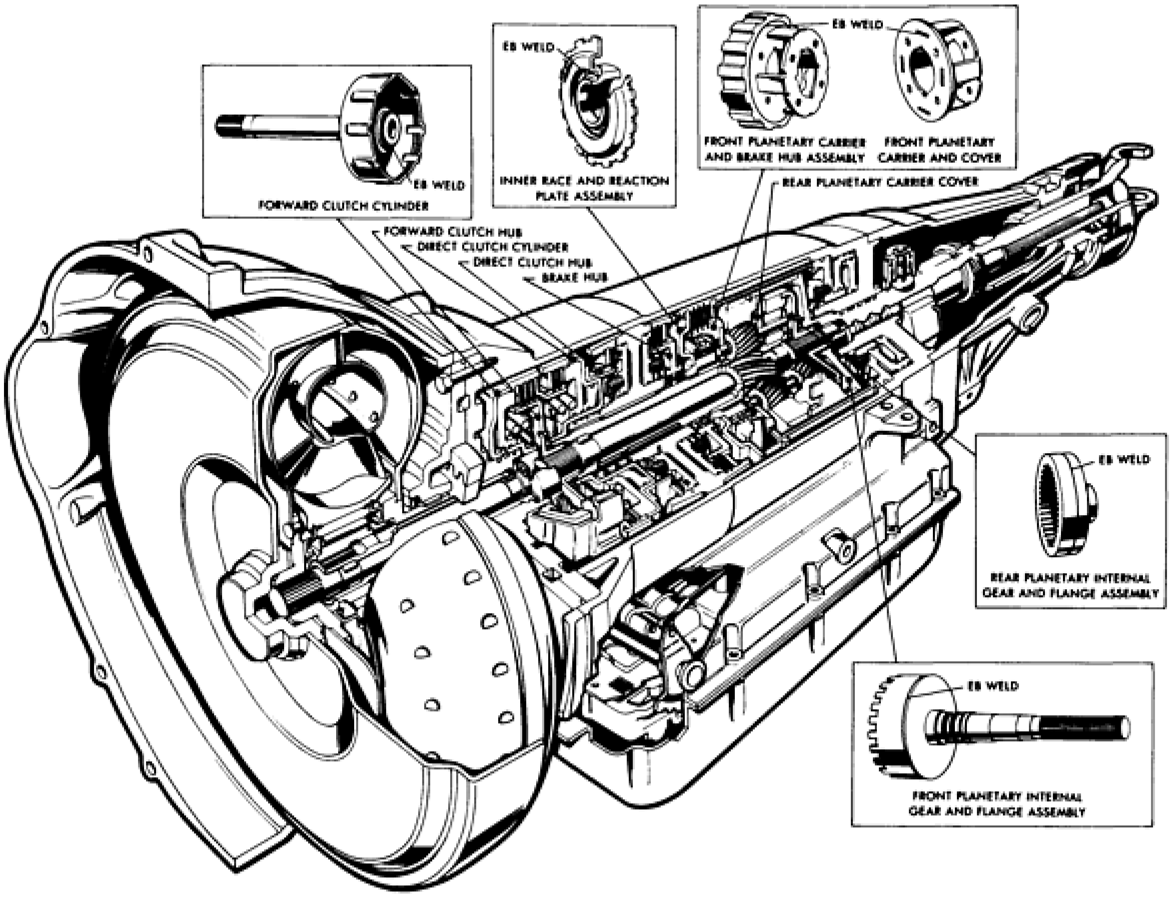


Fig. 1 Borg-Warner Model 12 automatic transmission. EB, electron beam

ATF Issues. The main method of wear protection relative to fluid maintenance is to maintain low levels of solid particulate contamination in the fluid reservoirs. Imbedded magnets, pump inlet filter screens, external cooling systems, and regularly defined fluid replacement schedules assure the highest performance levels in the automatic transmission. Because ATF is used to provide a heat rejection path, hydraulic power supply, and lubrication, many additives are used to serve the multipurpose functions for the fluid. Some applications result in up to a 10 vol% additive content in ATF.

Much literature has been dedicated to wear resistance and the related properties of ATF (Ref 1, 2). Specific benefits and detriments of the additives used to enhance ATF operating characteristics must be closely evaluated in the early development stages of a new product. For example, the compatibility of the ATF to the insulation systems of electrical components should be confirmed. This is typically done through the use of hydrolytic stability testing for the insulation system in short-term, high-temperature test programs.

Clutch bands and plates are fabricated from steel stampings with a variety of material coatings that provide the friction interface. In general, low-energy applications (~820 W/m^2) utilize paper-based or organic friction products, whereas high-energy applications (~1600 W/m^2) utilize bronze-base or semimetallic friction products. The performance

of both paper and semimetallic products, when properly applied and adjusted, has been observed to be acceptable over the full vehicle operating life (193,100 km, or 120,000 miles) without replacement. The normal wear associated with this type of application is 50 μm (2 mils) per band apply or clutch plate. For example, a five-plate clutch pack can be expected to wear a total of 250 μm (10 mils). The normal wear that is characteristic of a paper-lined clutch plate for 193,100 km (120,000 miles) of vehicle service is shown in Fig. 2. The distribution of this acceptable level of wear is critical to the overall success of the friction product application.

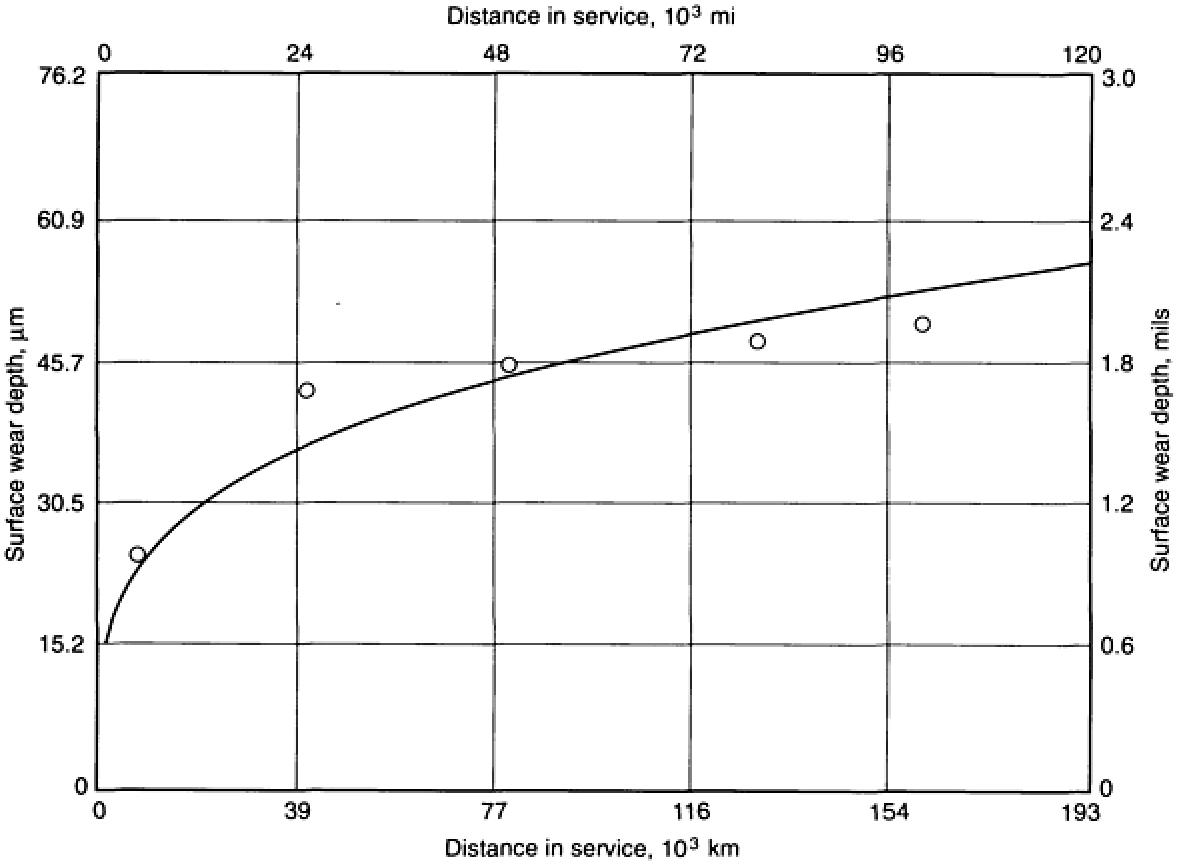


Fig. 2 Paper clutch material wear characteristic curve

Much work has been done in the design community to assure a high level of evenly distributed wear, thus providing a maximum life expectancy for the application. Paper and organic friction products characteristically have a more even wear distribution, compared with semimetallic and full-metallic clutch plates, because of the lack of resilience or compressibility of the metallic-based products. The basic elements for the successful application of friction products in automatic transmissions are provided in Ref 3.

With the increased use of electronically controlled transmissions, optimized shift strategies can be easily implemented to provide low-energy shift requirements throughout the entire engine speed and torque schedule. The engine management function typically allows a slight reduction in spark and fuel injection prior to the clutch plate or clutch band application. Once the shift has been completed in the transmission, the engine firing schedule is returned to normal. Typical shift times in the transmission are held below 0.7 s. The amount of engine retardation and elapsed shift time for optimal shifting are tailored to each application. The engine spark and fuel injection inputs can be reduced by up to 20% in some vehicles, with little noticeable effect to the driver.

Proper management of the hydraulic shift system also contributes to maintaining the lowest possible levels of wear. Some aspects of the hydraulic control system design for automatic transmissions are provided in Ref 4.

Clutch band and plate wear take the form of surface burnishing and polishing. High spots and surface imperfections are usually worn away very rapidly in early clutch applications, forming a wear-resistant surface with excellent wear-

resistance properties. In the case of paper and organic materials, compression set usually makes up the balance of the total estimated wear depth.

There are many variables, beyond energy dissipation, that must be considered when selecting the proper friction material for an application. Generally, material characteristics such as the rate of thermal conductivity, resilience, and fluid absorption play key roles. With respect to operating conditions, the presence or absence of lubrication fluid is the most critical, followed closely by a thorough understanding of the rate of applied load, engagement cycle time, and overall capacity objectives.

A review of the performance of automatic transmissions being provided to the market indicates that the optimization of design and manufacturing characteristics has been successfully achieved, resulting in high performance levels and satisfactory durability and reliability.

Manual Transmission Components

A typical manual transmission being produced in the early 1990s is shown in Fig. 3. The major source of wear varies, depending on the type of fluid fill used. Table 1 illustrates the variety of fluids used in manual transmissions currently in production.

Table 1 Various manual transmission fluids

Transmission	Manufacturer/application	Lubricant used
4-speed	GM/MC9	DEXRON II
4-speed	GM/M20	SAE-80W-90
5-speed	Getrag/MG290	SAE-05W-30
5-speed	Mitsubishi/Ford	SAE-80W-85
5-speed	Mazda/Ford	MERCON

Note: Fluid types based on review of 1990 GM and Ford light truck service manuals

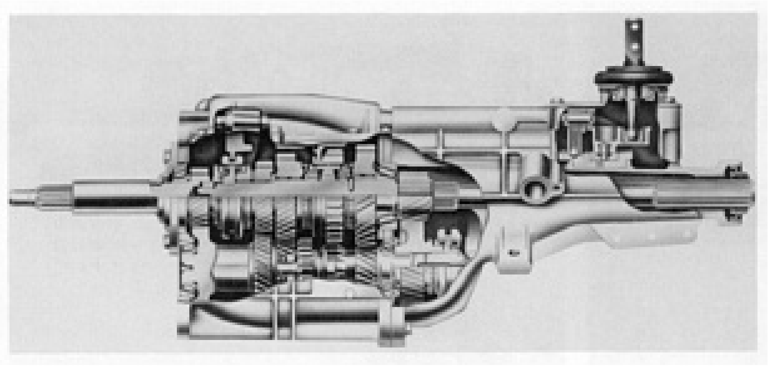


Fig. 3 Borg-Warner T-5 manual transmission

Engine oil and gear lubrication units generally experience little or no observable wear in the normal vehicle operating life. In this type of lubrication environment, gear mesh and bearing performance is generally superior to that of ATF-filled assemblies. An area of wear that is considered noteworthy in the gear lubrication equipped units is blocker ring tooth profiles and come synchronizer friction interfaces.

Synchronizer Wear. Examination of manual transmissions after extensive vehicle mileage accumulation often reveals some material loss on blocker ring engaging teeth. The wear takes the form of tooth edge rounding and flattening, which results from forced engagement under nonsynchronized conditions within the synchronizer assembly. This phenomenon is commonly referred to as clash. Detailed discussions of the symptoms and effects of clash, along with design practices to minimize its effects, are provided in Ref 5.

With respect to the come clutch friction interface, surface wear is a direct function of the energy dissipation when lubrication flow and availability are not compromised (as in the automatic transmission discussion). The same material selection rules apply relative to energy dissipation levels. Low energy requirements are best served by paper- and carbon-base products. High-energy applications utilize semimetallic or bronze-base materials. It is generally believed that cone geometry has increased sensitivity to uneven wear distribution, when compared to clutch band and clutch plate designs. This may be true, because of difficulties in maintaining proper cone-element alignments. Cone clutch performance for gear lubrication units in cold-weather applications is noticeably degraded, when compared to an equivalent ATF application. High-viscosity oils do not evacuate from the cone interface as readily, resulting in lower-than-anticipated dynamic friction values at the interface. This results in higher manual effort shifts and potentially accelerated blocker ring wear.

Cluster Gear Wear. In ATF-filled units, it has been observed that some cluster gear interface or mesh area wear would be considered normal. Typically, total wear of 125 μm (5 mils) would be permissible between the interfacing gears. The distribution of this amount is not considered to be critical, although it is obviously undesirable to accumulate the entire wear allowance in one gear set. The wear takes the form of surface polishing and some material loss at the tooth involute form (TIF). Detailed discussions relative to the mechanics of gear mesh wear are provided in Ref 6. The main preventative methods used to enhance wear characteristics for gears are to properly select base material and surface finish and to optimize gear design to provide the best pitch line velocity for a given torque capacity requirement.

High pitch line velocities can produce high localized heat zones in the contact area, because of friction. Therefore, reducing the load and speed conditions at the gear TIF improves wear characteristics. With regard to optimized gear design, some general sizing rules have proven successful over the years. Figure 4 illustrates the relationship between manual transmission torque capacity and center distance.

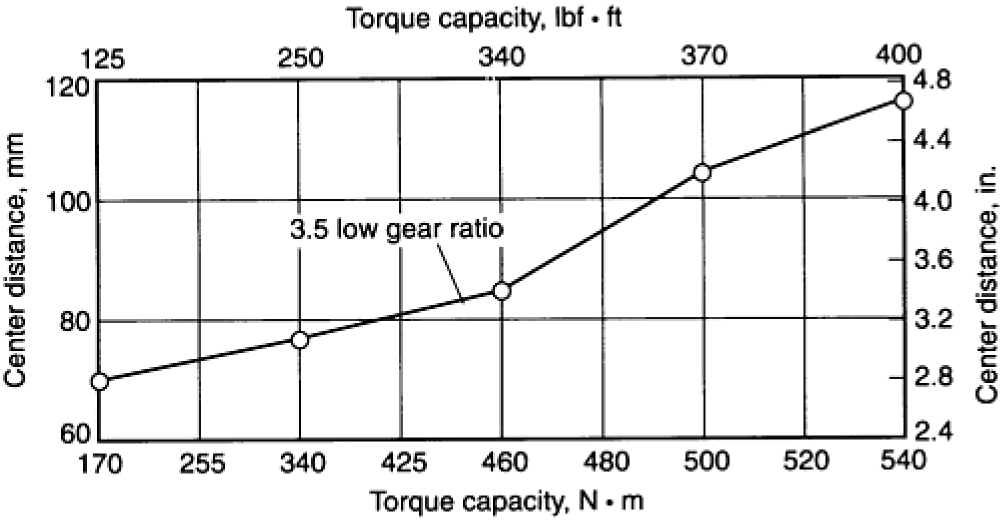


Fig. 4 Transmission torque capacity versus center distance

Significant departures from this generic curve have been achieved and are in production. However, large incremental improvements in torque capacity for no change in center distance are difficult to achieve and usually involve more exotic gear material selection, as well as additional surface finish and hardening processes. It should also be noted that the resulting ratings obtained from an individual line of the torque capacity graph are directly proportional to the depth of the first-gear ratio. Higher numerical first-gear ratios result in increasing the center distance requirement for a specified torque capacity objective.

Tapered roller bearings are widely used in current designs for manual transmissions. To a transmission designer, a significant source of wear involves the loss of bearing preload in the transmission, resulting in increased endplay and, consequently, revised gear mesh conditions. Figure 5 depicts the change in preload over the useful life of the transmission. Normal transmission wear can result in up to 250 μm (10 mils) of endplay on a unit that may have been originally assembled to a preload of from 50 to 75 μm (2 to 3 mils).

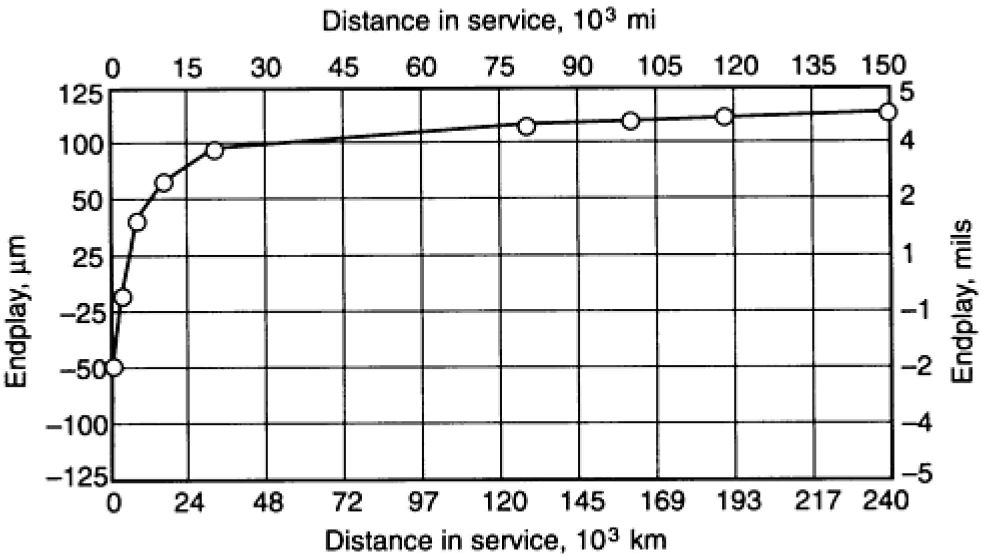


Fig. 5 Characteristic curve for change in assembly preload

There is some debate as to the root cause of the loss of preload. An acceptable premise for the wear may be that the tapered roller bearing elements, namely the outer bearing race, has reseated itself in a position determined by the dynamic action of the transmission during normal operation. Sensitivity studies have been performed, in which excessive (exceeding 250 μm, or 10 mils) endplay units have been built and subjected to various long-term durability tests. Test results indicate that the transmission can maintain satisfactory in-vehicle performance with no appreciable added wear.

Transfer Cases

Although the function and duty cycle for transfer cases vary significantly, their normal wear attributes are quite similar. Part-time units (Fig. 6), like those found in many popular off-road vehicles and light-duty trucks, exhibit chain and sprocket wear that is similar to those units that operate on a full-time basis (Fig. 7).

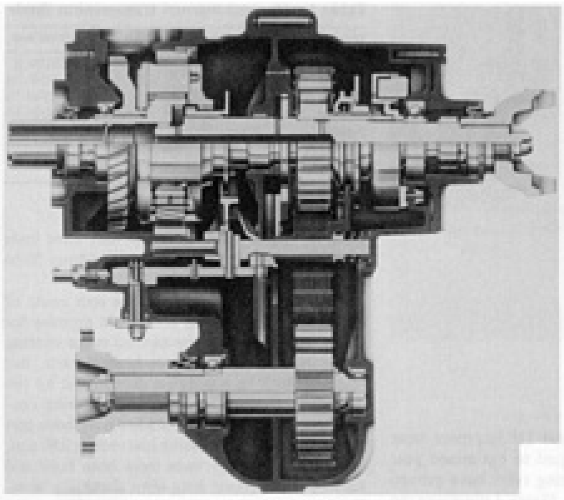


Fig. 6 Borg-Warner part-time transfer case

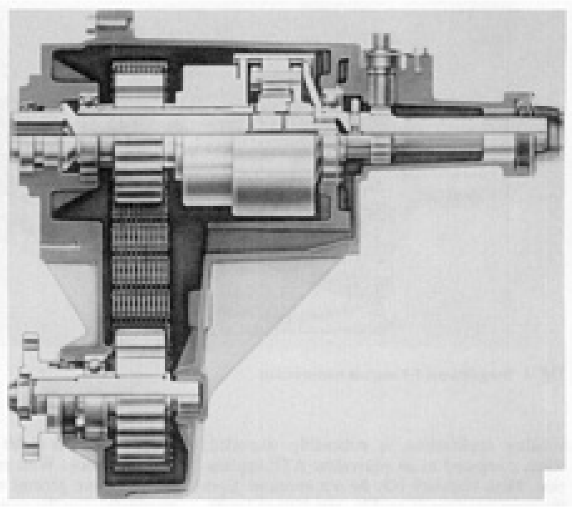


Fig. 7 Borg-Warner full-time transfer case

Sprocket and Chain Wear. The wear observed on the sprockets is a direct function of the chain mesh pattern. Many sprockets utilize sintered powdered metal technology to form the sprocket tooth. The normal wear for this tooth includes a depth from 125 to 250 μm (5 to 10 mils) on the tooth flank in a pattern formed from the hardened chain links. Local pitting zones are not unusual with long life or with heavy usage rate units that have interconnected pitting zones, forming a solid line across the tooth profile.

Some applications use wrought steel or forged sprockets, for which the wear characteristics are somewhat different. The drive sprocket wear appears as a series of depressions that are usually closely aligned with the chain-link construction. The driven sprocket wear takes the form of smearing across the chain sprocket mesh line. Drive sprocket wear is always heavier than that of the drive sprocket.

The wear of the chain has its own unique characteristics. The expression for chain wear is usually termed stretch. The maximum allowable stretch cannot exceed the length required to ensure that the chain will remain on the sprocket during operation. The point at which the chain can leave the sprocket tooth form can be described in terms of percent of pitch elongation. Chains that elongate beyond 25% of pitch are considered failures. The typical expected chain elongation for properly lubricated chain running on specified sprocket tooth forms is shown in Fig. 8. Experience with properly sized and lubricated full-time units has shown that normal expected chain stretch, in terms of percent of pitch, is on the order of 5 to 10%.

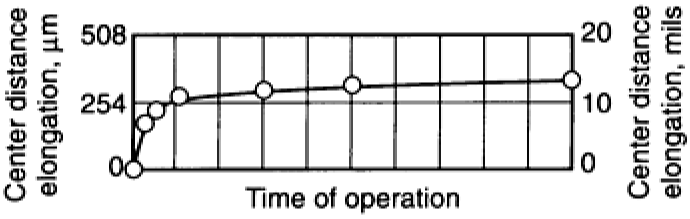


Fig. 8 Typical wear for chain applications

Several chain-link construction methods are in use today. Figure 9 illustrates two of the available configurations.

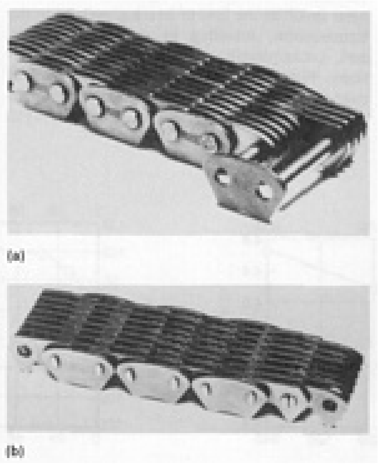


Fig. 9 Chain design examples. (a) Round-pin design. (b) Rocker-joint design

For transfer cases, the basic approach taken to accommodate the sprocket tooth form wear and chain stretch potential is to design an adequate chain operating envelope into the surrounding castings to ensure proper lubrication and minimize the development of churning effects. When sizing a chain for a given application in a transfer case, the maximum available engine torque and vehicle weight ratings are used to determine the chain pull load. Equations 1 and 2 represent the maximum available engine torque and skid torque, respectively.

$$T_{\text{stall(automatic)}} = E_t g_a t_s \quad \text{(Eq 1)}$$

where E_t, the maximum available engine torque, is 300 lbf · ft; g_a, the highest numerical gear ratio, is 2.7; and t_s, the torque converter stall ratio, is 1.8.

$$T_{sf} = \frac{\mu R_r F_{gawr}}{A_r \, 12 E_p} \quad \text{(Eq 2)}$$

where μ, the coefficient of friction, is 0.8; R_r, the tire rolling radius (dynamic), is 11.93 in.; F_{gawr}, the front axle gear wear rating, is 2926 lb; A_r, the axle ratio, is 3.62; and E_p, the power-train efficiency, is 0.85.

The application torque through the chain is directly related to the chain pull force and, consequently, to the degree of chain stretch observed over the useful life of the product. Figure 10 shows actual field measurements made on chains in service in an all-wheel drive application for units ranging from very low mileage accumulation to over 160,900 km (100,000) miles) of service. Curves of this type are developed for part-time units, as well. However, because of the variability associated with the degree of full-load chain service, wear data are not as easily interpreted.

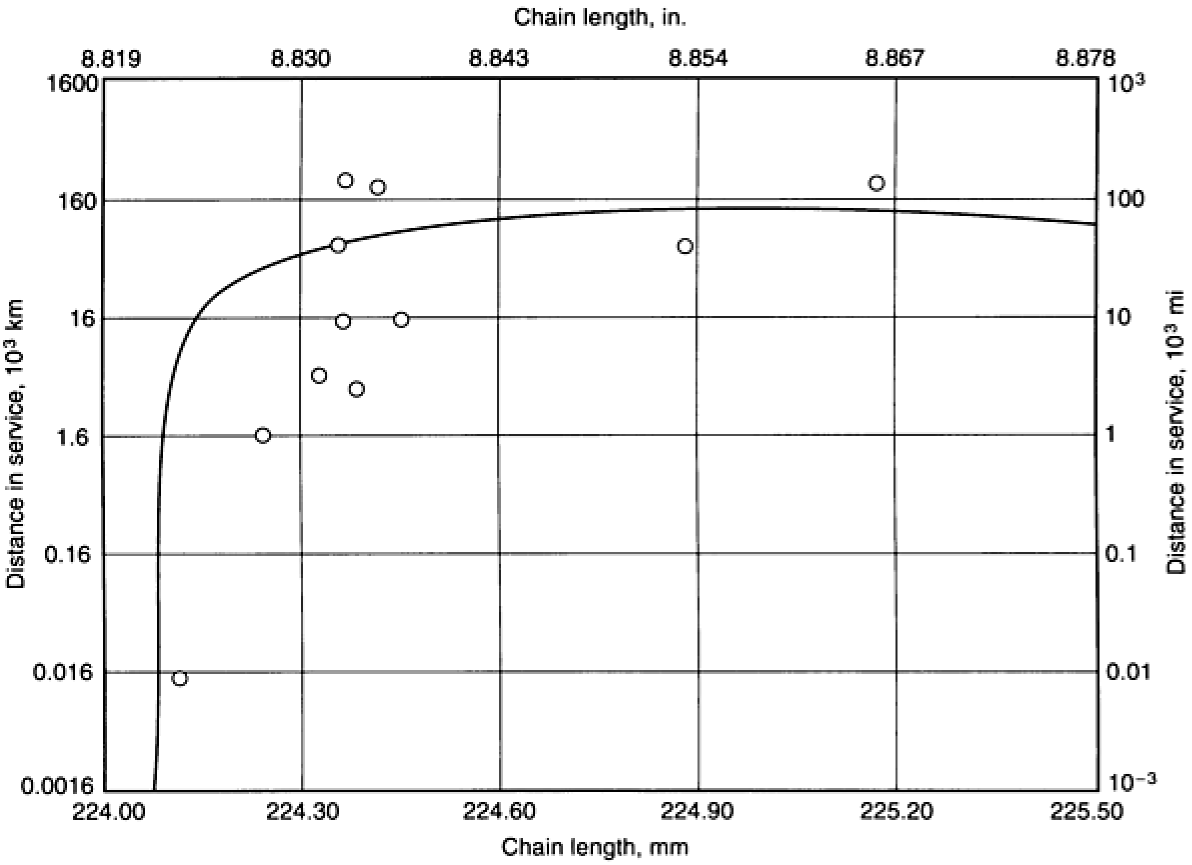


Fig. 10 Chain stretch field data

Another important wear characteristic observed in part-time transfer cases equipped with speed-reducing gear sets and internal shift systems is thrust surface wear. Because of the part-time nature of the transfer case operation, the potential for combined high-speed differences and axial loading for internal components is high. The two major wear areas can be found in the shift system fork pad and the helical reduction gear set thrust surface areas.

Thrust Surface Wear. Many transfer cases manufactured today use cast aluminum or cast iron shift forks with polymer inserts for dog-clutch movement. After many miles of continuous duty, the polymer insert material may wear to the point that the dog-clutch positioning is compromised and the vehicle does not maintain drive. This condition is precipitated by small axial loads or thrust loads between the rotating dog-clutch flange and the stationary polymer insert pad.

Material selection for the insert pad must consider the degree of thrust load applied, in conjunction with maximum speed differential and anticipated duty cycle. The load and speed relationship to wear has been well investigated and documented in terms of polymer compositions. Material suppliers are generally able to provide load and velocity capability for a material, *PV,* as well as wear factors, *K,* which are usually defined as the ratio of volumetric material reduction, *W,* to applied force, *F,* velocity, *V,* and elapsed time, *T.*

$$K = \frac{W}{FVT} \quad \textbf{(Eq 3)}$$

In addition to superior *PV* performance, polymeric materials should have high impact resistance in this type of application, because of the cyclical nature of the thrust loading. This characteristic separates the fork pad wear thrust materials from those used in reduction gear sets. In the gear set applications, the speed difference ratio between two thrust members is held constant, because of the nature of the gear set design. The forces developed between thrust surfaces in reduction gear sets is directly related to the gear set geometry (helix angle) and the axial power-train compliance.

Although the gear set geometry effect is readily available through calculation, the power-train axial compliance is a much more difficult parameter to quantify. When high thrust loads are present and the application variability is significant, thrust washers are provided on all interfacing surfaces. The most widely used and successful thrust washer material is VESPEL, a DuPont product. It is a polyimide resin that can be available in direct-formed shapes or machined from solid stock.

Engine Components

This section reviews the interesting wear characteristics for engine turbochargers. Significant material is provided in the article "Friction and Wear of Internal Combustion Engine Parts," relative to wear.

The durability and reliability of turbochargers is often underestimated in the North American marketplace. Turbochargers can provide 99.5% reliability for 160,900 km (100,000 miles) of vehicle service. For the turbocharger designer, the key challenge, with respect to wear prevention, revolves around a thorough understanding of sealing and cooling in an environment where shaft speeds can reach 150,000 rev/min and maximum gas temperatures can be 705 °C (1300 °F). A cross section of a typical turbocharger design for an automotive engine application is shown in Fig. 11.

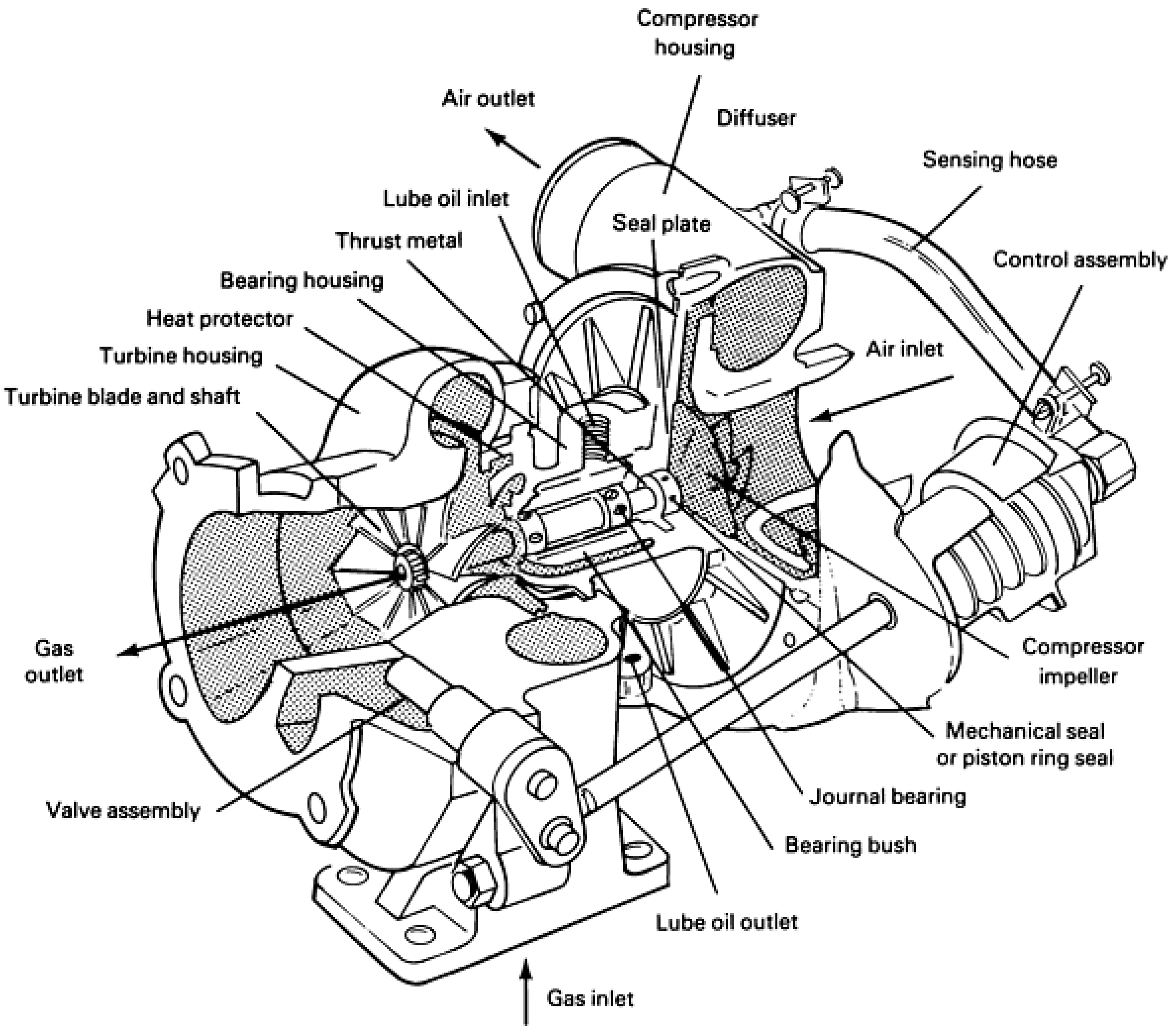


Fig. 11 Warner-Ishi RH series turbocharger

As is the case for most mechanical systems, rotating members must be provided with adequate lubrication to ensure long-term operation. The turbocharger uses an intricate lubrication strategy to deliver controlled quantities of oil to mainshaft journal bearings and bushings, in order to minimize wear and take advantage of hydrodynamic and aerodynamic sealing technologies. In early turbocharger designs (before the use of fuel-injection systems), the compressor housing/turbine

housing interface design utilized a mechanical seal with a carbon-base material and a combination thrust washer seal plate that provided controlled deformation and material transfer to account for the very high *PV* conditions created by the normal operation of the turbocharger. Unlike seals found in normal power-train applications, turbocharger mechanical seals were most efficient and effective at very high operating speeds and pressures. At low speeds, the seals were inefficient and accounted for high drag losses in the assemblies. Many of the turbochargers used with fuel-injection-equipment engines use metallic piston ring seal designs. These designs emphasize improved aerodynamic features to lower the "blow-by" potential in the piston seal location.

With respect to the wear and deformation that occurs at the thrust plate, a review of the general maintenance procedures for turbochargers indicates that the expected wear on the thrust washer seal plate for an "end-of-life" unit would range from 25 to 38 μm (1 to 1.5 mils) of endplay. The most significant wear contributions have been observed in the gasoline-engine-equipped systems, where very high operating temperatures are present and external cooling systems may not be readily available. Under these circumstances, the mainshaft bearing temperatures can become elevated to the point that evaporative lubrication residue (coking) can significantly reduce the efficiency of the hydrodynamic bearing system. Diesel-engine-mounted systems have traditionally operated at much lower temperatures and have improved performance.

Some designers are in the process of evaluating fluoropolymer sealing systems. These systems are required to function in the same high-speed, high-temperature environment as the mechanical seals. Because fluoropolymers have good thermal degradation resistance at high temperatures and very good self-lubricating characteristics, their use in this type of application has promise.

Wheel End Components

Part-time four-wheel drive vehicles require the ability to engage and disengage the front drive axle during vehicle operating mode changes from two- to four-wheel drive. For many years, this function has been provided by automatically engaging wheel end hubs. Figure 12 illustrates a typical torque-activated automatic locking hub.

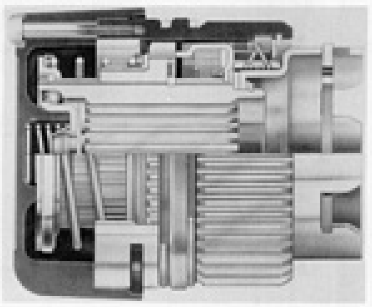


Fig. 12 Borg-Warner automotive 17-01 automatic locking hub

Because the principle of operation for the locking hub requires that a near-synchronous shift potential be provided by some external torque source, and because the locking hub has no internal synchronizer capacity, the preventative wear and long-term durability objectives are based on providing very high wear resistance for small, thin cross-section gears and cam surfaces. The lack of a renewable lubrication supply and the difficulties in providing long-term sealing features under the extremes of temperature and vibration result in careful material selection, exotic lubrication strategies, and inherent design features that will allow nonsynchronized gear mesh to take place without much resulting wear.

Typically, a dog-clutch arrangement is employed to provide the physical connection between the wheel end and the axle shaft spindle. The axial displacement of the dog clutch is accomplished with the positive displacement features of cylindrical cam surfaces under rotation. The dog-clutch/cam surface interface depth of wear is usually not allowed to exceed 125 μm (5 mils) over the life of the product. The wear tolerance is usually taken up in the cam follower tips and flanks and in the inner cam surface. Springs are used to assist in the axial displacement of the dog clutch during the engagement cycle. Because there is typically no internal synchronizer feature, the spring duty cycle is a direct function of the success of penetration of the dog clutch into the overspeeding receiving gear.

Special tooth chamfers are employed in an effort to reduce tooth edge or butt wear during the penetration. The chamfer surface may be expected to wear on the order of 380 to 635 μm (15 to 25 mils) for the full service life. The degree of wear is directly related to the multiple high-speed engagement duty cycle. Engagement systems that do not provide some gear tooth form chamfer in the dog clutch often show heavier wear and undesirable shift quality as a result of repeated tooth edge engagements. These types of systems generally have a higher degree of brinelling in the tooth edge area, to the point that the effective tooth form is wider, thereby further reducing the likelihood for successful dog-clutch penetrations.

References

1. M.L. Haviland and J.J. Rodgers, "Friction Characteristics of Automatic Transmission Fluids as Related to Transmission Operation," Paper 60AM6A-1, presented at the American Society of Lubrication Engineers Annual Meeting, Apr 1960
2. W.K. Kite and E.F. Koenig, "Evaluating Automatic Transmission Fluids," Paper 13R, presented at the Society of Automotive Engineers Annual Meeting (Detroit, MI), Jan 1959
3. Ratio Changes--Friction Elements, Section III, *SAE Design Practices--Passenger Car Automatic Transmissions,* Vol 5, Society of Automotive Engineers, p 79-138
4. Ratio Changes--Friction Elements, Section VIII, *SAE Design Practices--Passenger Car Automatic Transmissions,* Vol 5, Society of Automotive Engineers, p 313-339
5. R.J. Socin and L.K. Walters, "Manual Transmission Synchronizers," SAE Paper 680008, Society of Automotive Engineers
6. The Design, Manufacture, and Application of Gears, *Gear Handbook,* 1st ed., D.W. Dudley, Ed., McGraw-Hill, 1962, p 14-43

Friction and Wear of Automotive Brakes

Arnold E. Anderson, Consultant

Introduction

AUTOMOTIVE BRAKE LININGS, as a group, include those used for passenger car and light truck drum and disk brakes, heavy truck and bus drum brakes, and off-road vehicle drum, band, and disk brakes. This article addresses all of these friction materials that nominally operate dry. Wet brake linings, that is, those that are oil immersed, are beyond the scope of this discussion.

Brake friction and wear characteristics are sensitive to brake design and lining composition effects, as well as to temperature, rubbing speed, pressure, and prior usage history. Wear of the brake drum or disk is often more sensitive to road dust and local rust debris than to friction material composition.

The wear and frictional behavior of these material pairs is amazingly complex, and is characterized by nonsteady-state high-temperature and high-pressure processes. Contact asperities reach flash temperatures ranging from 1000 to 1100 °C (1830 to 2010 °F) in about 1 ms, and then cool as other contacts become active. To avoid thermoelastic instabilities, heterogeneous and often highly anisotropic friction materials have been developed. Because meaningful patent protection is difficult to obtain, friction materials tend to be highly proprietary formulations.

Poorly understood tribological mechanisms and obscure compositions preclude the ability to present a classical, scientific approach to brake lining friction and wear. However, a survey of brake performance effects is provided, along with plausible explanations or underlying basic mechanisms. Material selection criteria for different brake designs and intended usages are also discussed.

Automotive Brake Linings

Automotive brakes include drum and disk brakes for passenger car, light truck, bus, heavy truck, and off-road vehicles. Drum brakes predominantly use internal expanding shoes with brake linings that load the majority (typically 50 to 70%) of the drum rubbing surface. Most automotive disk brakes use shoes that load a much smaller portion (from 7 to 25%) of the disk rubbing surface.

Disk brakes faster cooling, with their larger exposed surface areas and better cooling geometry, but are more vulnerable to either liquid or solid particulate contamination. Because of cooling, contamination, and other basic design issues, front disk brakes and rear drum brakes are commonly used. The friction material is called a brake lining, whether it is used on drum or disk brakes. Disk brake linings are also called pads, drum brake linings are known as segments or strips, and heavy truck drum brake linings are called blocks.

Brake linings are the sacrificial wearing element and the primary determinant of frictional behavior. However, the brake drum or disk material, usually a gray cast iron, also can affect brake lining friction and wear. Automotive friction materials have been formulated for about 90 years. There are three broad classifications of friction materials: organic, metallic, and carbon.

Organic friction materials were exclusively composites of one or more binder resins, chrysotile asbestos fiber, and a blend of additives that modified friction, improved wear, reduced cost, aided processing, changed color, and so on. Currently, there are three subclasses of these resin-bonded friction materials:

- Asbestos, also called organic brake linings
- Nonasbestos organic (NAO), using a variety of fibers
- Semimetallic, or resin-bonded metallic (semimet)

Semimet linings are separate from the other NAO linings because they have a restricted composition range, with unique friction and wear properties. Semimet linings comprise about 65 wt% total iron content, 10 to 25% steel wool, and the remainder is commonly a porous iron powder. Graphite, usually synthetic, is about 15 wt%. A heat-resistant phenolic-type binder is about 10%.

NAO brake lining formulations are rapidly evolving, and may eventually develop into several subgroups. Currently, these formulations designate what they do not contain, such as nonasbestos nonmetallic. Over 1200 different fibers, acicular materials, and other reinforcing agents have been tested to date. Because most NAO materials use a blend of different fibers and other reinforcing agents, the number of potential combinations is staggering. The most commonly used fibers are glass (including chopped glass fiber, mineral wool, and many proprietary species), metal, ceramic, para-aramid (Kevlar), cellulosic, and other organic forms. Acicular materials, such as wollastonite, and plate-like materials, such as attapulgite, are also used as reinforcing agents in NAO linings.

The binder resins used in organic friction materials are typically thermosetting polymers, often a two-step phenolic type. Both liquid and powder resin forms are used. Rubber and cashew (derived from an oil in the nut shell) binder resins also are used in brake linings.

For good wear properties, cured brake lining resins should be tough, heat-resistant, and have predictable thermal degradation behavior. Binder resins that readily break down into oily decomposition products are suited only for light-duty drum brake linings. Their oily decomposition products combine with solid wear particulates to form a surface film material that enhances low-temperature wear life. However, such linings would likely self-lubricate (fade), wear excessively, and/or slough away in hard service. A binder resin requires thermal stability for the most severe brake usage anticipated.

Resins that form a stable char at the rubbing surface are better for higher-severity applications, but may provide a lower friction level. When used in heavy-duty friction materials, they do not provide "oily" breakdown products at the rubbing interface to retain wear particulate at low-severity conditions. Consequently, low-temperature wear life can be poor. Organic additives can be used to enhance light-duty wear. Cashew particles, at 7 to 12 wt%, are often used to provide such wear debris reutilization.

More-stable organic compounds can be used to enhance wear in higher-severity applications. Para-aramid, at levels from 2 to 3 wt%, has been shown to effectively improve wear life under more stringent usage conditions. Resins and other organic brake lining constituents that char or form oily degradation products, depending on usage conditions, may provide inconsistent friction and wear.

Wear properties of the friction material also may be enhanced by the appropriate choice of inorganic constituents. In general, lining additives can be used to improve strength, thermal expansion, heat absorption, friction, or other properties. The friction material rubbing interface can be thought of as a high-temperature, high-pressure, nonequilibrium thermal reactor. The effects of most friction material additives on brake performance cannot be predicted accurately because of insufficient fundamental knowledge. Consequently, most brake lining formulation is achieved through trial and error, coupled with prior experience and testing expertise. A formulation involves many compromises, and is beyond the scope of this article. The remaining two classes of friction materials have unique performance characteristics, but with limited automotive applications.

Metallic brake linings can be based on either copper or iron. Most are solid-state sintered, often with inorganic additives to improve performance. These have been developed for very high power input densities. For example, solid-state-sintered bronze and mullite linings are used in race car and high-speed railroad brakes. Sintered iron with graphite is used on some heavy-duty brakes, both disk and drum, as well as on a few production passenger car drum brakes.

Another metallic friction material is cast iron. Although this venerable material is used on some old railroad tread brakes, no new automotive applications are known to use cast iron as the brake lining. However, it is the predominant countersurface material used for automotive drum and disk brakes. Therefore, it remains a friction element, although not a brake lining.

Carbon-based brake linings, such as carbon-carbon, were developed for military and commercial aircraft disk brakes. Some are now used on racing cars where weight is critical, performance is demanding, and cost is secondary. Carbon-carbon friction materials are made from carbon fiber (also called graphite fiber) that is bonded with amorphous carbon. Organic resins are either baked at high temperatures or a chemical vapor deposition process is applied to generate this amorphous carbon binder. After further processing, the resultant friction material is essentially pure carbon with a very low porosity. A 2000 °C (3630 °F) operating temperature and a high specific heat property permit carbon-carbon materials to provide weight savings of nearly 85% when replacing cast iron in suitable applications.

Brake Lining Wear

Brake lining wear life frequently is measured in imprecise terms, often as some number of miles or kilometers in service. Brake adjustments for wear, when required, usually are made on the basis of either usage time or distance traveled, utilizing prior experience. Such rough measures are acceptable for many users, but unacceptable to most researchers and engineers. In this article, the brake lining specific wear rate denotes the mass worn (in mg) per unit of frictional work (in MJ, or about 0.37 hp · h). Good brake linings have values from 40 to 100 mg/MJ (0.003 to 0.008 in.^{3}/hp · h) at low temperatures. High usage temperatures can increase these wear rates by one to three orders of magnitude.

Substantial testing may be required for wear documentation, because brake lining wear rates change greatly with temperature, and brake usage varies enormously in automotive service. This testing can be reduced by understanding the mechanisms of wear, which can be investigated through model studies.

Break-In Wear. During the necessary burnishing, or wearing in, of brakes, both the brake drum or disk and the friction material wear at a temporarily high rate. Brake lining wear rates are usually stabilized by the time the major brake drum or disk asperities (those due to machining) have been rounded. Normally, the lining break-in wear volume is small, compared with the total available wear volume. Many break-in factors exist. The more important of these have been reported by Rhee and Thesier (Ref 1).

Brake Torque Effects on Wear. Most organic linings show essentially no wear variation with applied force, but semimet linings almost invariably show rates of specific wear that are several times higher than normal at very low loads. This high lining specific wear rate produces a characteristic fine red wear particulate (mostly ferric oxide) and results in a greatly altered cast iron countersurface wear rate (described more fully later). With this one major exception, brake lining wear rate can be considered independent of applied load.

Brake Rubbing Speed Effects. Organic brake linings show little variation in their specific wear rate with braking speed, when tested from low initial cast iron temperatures. There typically is a slight rise of wear rate below a rubbing speed of 2 m/s (6.5 ft/s). Brake asperity "flash" temperatures are known to vary primarily with speed, as has been described by Blok (Ref 2) and, more recently, by Lim and Ashby (Ref 3). Above 1 m/s (3.3 ft/s), asperity temperatures appear to range from 1000 to 1100 °C (1830 to 2010 °F). It is presumed that the variation of lining wear rate and the associated variation of friction level are related to this flash temperature transition. At very high rubbing speeds, the lining wear rate increases. This increase is greater when the initial brake drum or disk temperature is high. It is presumed that this speed effect is simply the result of higher interfacial temperatures. Model wear data, presented later, support this presumption.

Semimet friction materials also exhibit unique behavior with speed. At rubbing speeds below 2 m/s (6.5 ft/s), semimet lining wear rates also increase, but to a significantly greater extent than do the organic linings. Thereafter, the semimet brake linings provide a nearly constant specific wear rate with rubbing speed, until a transition condition is reached. Higher speeds then generate much higher lining wear rates (to 100 times, or more). During full brake dynamometer testing of semimet brake linings, it was found that four brake stops from 160 km/h (100 mph) produced as much lining wear as over 500 brake stops from 50, 65, 100, and 130 km/h (30, 40, 60, and 80 mph).

There are differences among commercial semimet brake lining formulations in terms of rubbing speed transition values. Higher transition speeds were found with a semimet lining that contained a small amount of para-aramid pulp. It is conjectured that this high-strength thermoplastic material provided enhanced near-surface brake lining strength and helped to prevent "friction welding" wear of the lining to the brake rotor.

Brake Temperature Effects. The variation of brake lining specific wear rate with brake drum or disk cast iron temperature is shown in Fig. 1 for four representative lining classes. The friction materials are divided into semimet (SM), light-duty (LD) organic, heavy-duty (HD) organic, and original equipment (OE) organic classes. Because asbestos and NAO materials have overlapping wear properties, they are not separated.

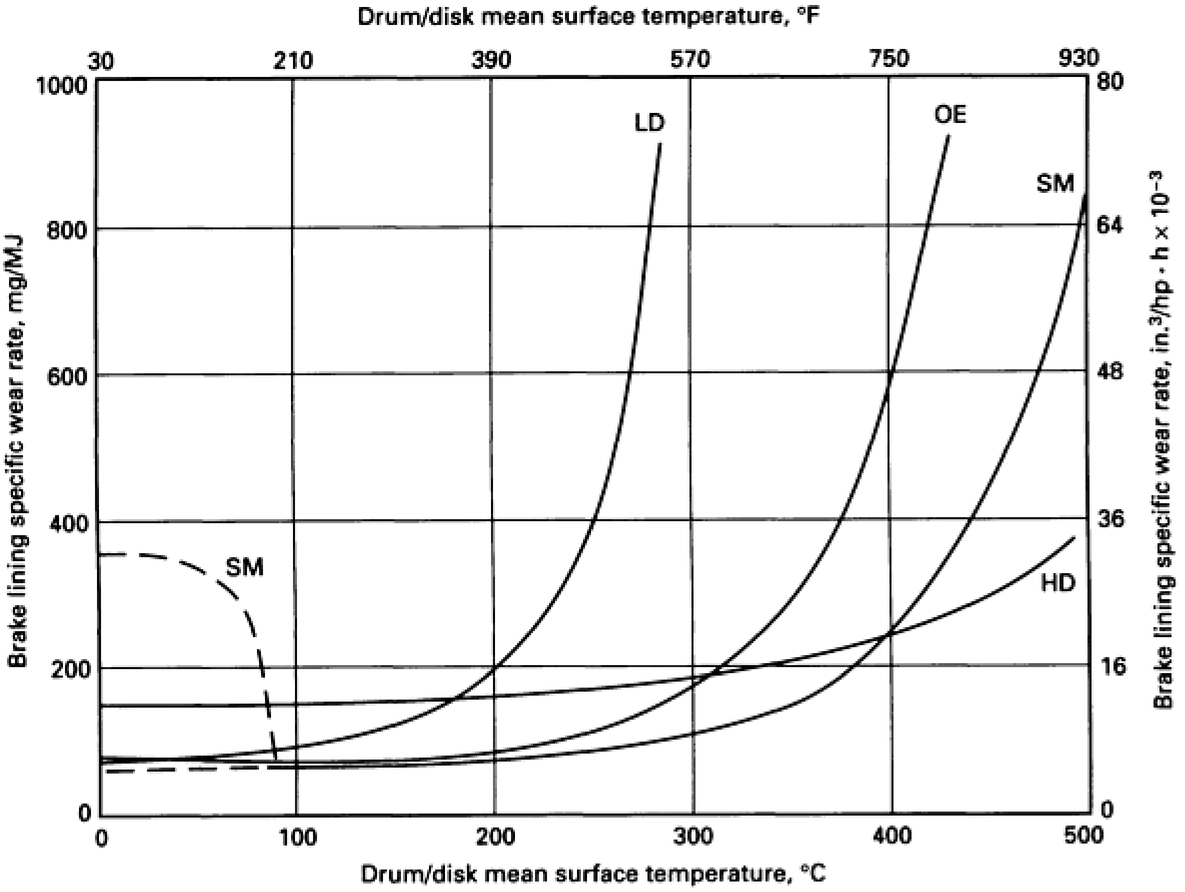


Fig. 1 Temperature effects on brake lining specific wear rates

Semimet linings also exhibit a typical behavior that requires some explanation. These friction materials establish a transfer layer to the cast iron during break-in and initial service. Until established, the semimet wear rates are several times the normal rate.

This cast iron conditioning process occurs faster with higher surface temperatures, higher rubbing speeds, and higher unit pressures. It also varies somewhat among the different semimet lining formulations, and even with differences of cast iron countersurface, such as texture, residual stress, and oxidation.

Once the cast iron surface is conditioned, it is believed that a back-and-forth transfer of material takes place. Material exchange between the cast iron transfer layer and the semimet lining surface helps to keep the measured lining wear rate low. No direct cast iron contact or wear occurs, except for incidental scoring that is due to particulate contamination, after the transfer layer is established. This is further described in the section on brake drum and disk wear.

Brake Usage Severity Effects. Figure 2 shows the wear life behavior for several classes of brake linings under different severities of usage. It can be seen that no friction material type is best for all usage conditions. Inexpensive aftermarket (AM) friction materials have acceptable wear lives under light usage conditions, but wear rapidly under more stringent conditions. Heavy-duty (HD) friction materials have a lower wear life variation with usage severity than most automotive materials, but are superior in wear behavior only for severe-duty usage. Sintered metallic friction materials also may have this wear characteristic, with low variations of wear rate at different usage conditions. Semimet (SM) brake linings are best in the middle range, but poor in very light duty conditions. Original equipment (OE) brake linings have a broad range of acceptable wear life. From Fig. 2, it should be clear that friction material wear life values are essentially meaningless unless usage conditions are specified. It is not possible to select the optimum friction material for wear without knowledge of the customer brake usage distribution.

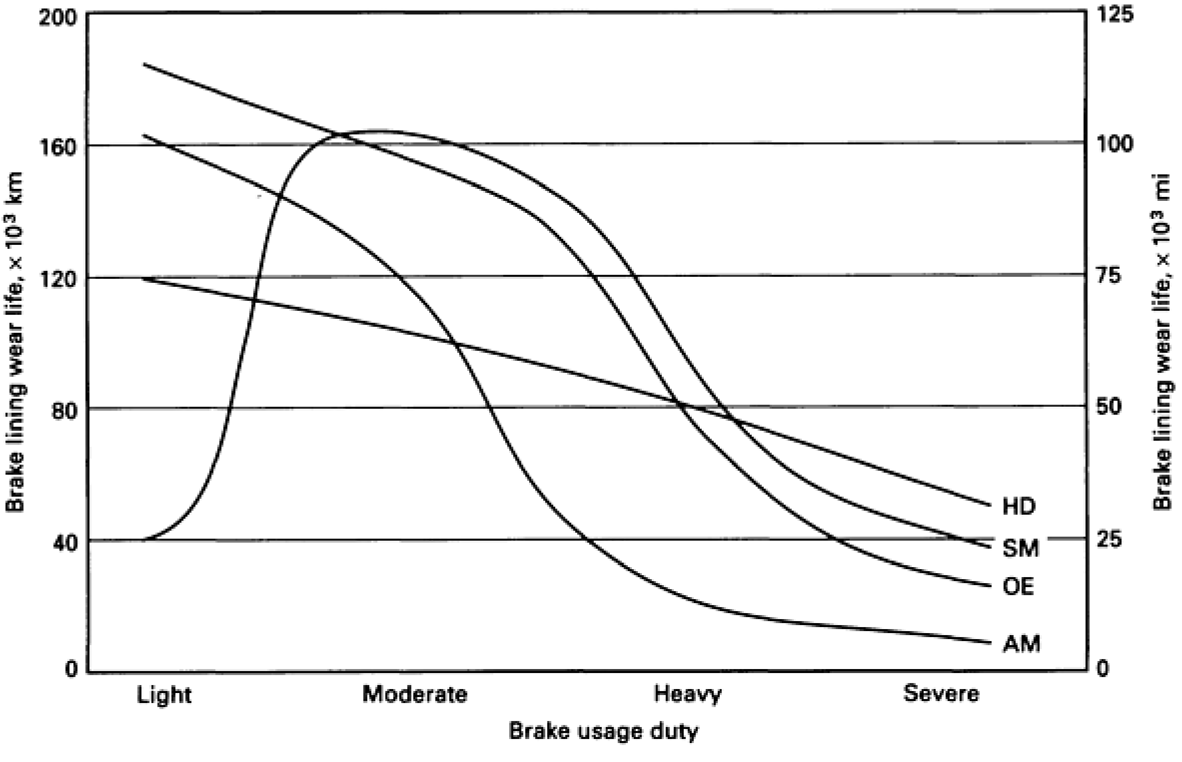


Fig. 2 Brake lining wear life versus usage severity

Brake lining Wear Modeling

Archard's equation is applicable to automotive brake systems only at very low rubbing speeds and component temperatures. Consequently, this linear relation between volume of wear and the product of sliding distance and applied load is not generally applicable for brake wear modeling. Forcing friction material wear data to fit polynomials of speed, rubbing distance, and load terms has been tried, but with very limited applicability.

Rhee (Ref 4) and others have related measurements on friction material wear with the brake cast iron temperature. These data indicated an essentially constant wear rate for low temperatures and a nearly exponential one at elevated temperatures.

Lining Cure Effect. Weintraub and Bernard (Ref 5) derived a model equation which showed that some pyrolytic gas chromatography (PGC) peak areas of a brake lining resin behaved in an orderly manner with curing time and temperatures. Simple laboratory-prepared formulations were used. They also showed that laboratory friction material wear rates for these simple brake linings were linearly related to these PGC peak areas. This work led to a simple, second-order Arrhenius-type model that was expressed as:

$$m = W + Kte^{(R/T)} \qquad \textbf{(Eq 1)}$$

where m is the mass loss (on a test of fixed frictional work), W is a wear constant, K is related to the Arrhenius collision coefficient, R is related to the Arrhenius activation energy, t is the brake lining cure time, and T is the brake lining cure temperature.

Additional studies by Weintraub, Anderson, and Gealer (Ref 6) showed that this wear relationship was applicable to additional friction material binder resin types in brake lining model formulations that approximated production linings.

Interfacial Temperature Effects. A similar expression was shown by Anderson (Ref 7) to be applicable to most brake lining wear data, but now relating the in-use friction material wear rate with the mean interface temperature by

$$V = A + Be^{(C/T)} \qquad \textbf{(Eq 2)}$$

where V is the specific wear rate (wear volume per unit frictional work), A is the low-temperature lining wear constant, B is a friction material wear constant, C is another friction material constant, and T is the average interface temperature during braking.

This equation has been quite effective in relating brake lining wear data with testing conditions for a broad range of materials, over most of the brake operating temperature range. Model formulations and production friction materials had full-brake wear behavior that was found to be well-characterized by this relationship. The production linings had several organic components, making the applicability of this equation surprising. Experimental values found for the constant C appear to have functional significance. For most friction materials, this activation-energy-related term has been determined experimentally to vary from 65 to 85 kJ/mol (16 to 20 kcal/mol). Such values might well be expected from bimolecular reactions in the wear process.

Because the constant term A dominates at low operating temperatures, it is easily determined experimentally. However, following some severe usage conditions, especially after high friction material soak temperatures, the value of A was found to increase and subsequently remain at this increased value. The increase appeared to be associated primarily with the friction material soak temperature, but also with some dependence on the time at temperature. It does not appear to be simply related to the mean interface temperature during braking. Thus, a better model expression would have the constant term replaced by one that includes prior thermal history. This new term also could include particulate contamination factors for an even more sophisticated model.

As the temperature increases, the thermally sensitive Arrhenius-type term increases in influence. The temperature required in the above wear equation is the mean rubbing interfacial temperature. This is not easily measured, but it can be calculated. The readily calculable interfacial temperature rise for the brake application is added to the measurable bulk cast iron temperature of the brake drum or disk. Values for B and C can then be determined from experimental wear data for temperatures, usually in the 150 to 350 °C (300 to 660 °F) range.

Whether this wear equation is technically correct or not is not at issue. It generally works well, and unique wear mechanisms appear to be involved when it does not. To a pragmatic experimentalist, this has been quite acceptable.

Brake Drum and Disk Wear

Gray cast iron is the dominant material for both brake drums and disk brake rotors. Brake cast iron typically has a type A graphite, with a pearlitic matrix and low ferrite and carbide content. Normal brake iron provides adequate damping and good resistance to thermal fracture in hard service.

The wear of either the cast iron or other countersurface is not readily determined by the Arrhenius-type model relationship. It appears that several mechanisms affect the specific wear rate of the cast iron. These include abrasive, adhesive, and oxidative terms. The abrasive terms include components from the abrasive content of the friction material, external contaminants, and abrasive particles that are "manufactured" at the rubbing interface. The latter can result, for example, from vitrification (firing) of clays at the hot rubbing surface.

Brake Lining Chemistry Effects. Some friction materials have constituents that are chemically active at the rubbing interface. Such brake linings have been found to be extremely responsive to changes in the cast iron chemical composition. Hatch (Ref 8) described the effect of titanium and rare earth oxide content on lining friction. Examples of large brake effectiveness differences with the same brake lining, but different cast irons, were shown. Regrettably, this article did not make clear that most friction materials do not provide any detectable difference in effectiveness, only the few that chemically interact with the cast iron.

With one such disk brake lining, the in-service cast iron wear life was increased by a factor of 50 when the titanium content was increased from 0.02 to 0.04%. The presence of small, hard particles in the cast iron from the titanium was credited for the improved wear resistance on this countersurface. Titanium content control can be particularly effective in disk brakes, if parasitic drag of the brake linings causes local wearing of the rotor. Cast iron machinability considerations limit the useful titanium content of 0.05%.

Graphite Morphology Effects. The graphite size and shape were found to affect the cast iron and lining wear rates for many brake linings. "Damped" gray cast iron, with its very large graphite flakes, has been used in some brake applications to decrease brake squeal. However, this iron is weaker, and tends to have poorer wear resistance than conventional brake iron. It also increases brake lining wear rates for some friction materials. This presumably results from the cutting action of sharp iron edges found around the large graphite flakes. Linings with hard resins and rigid matrixes were found to have lower wear rates with decreasing graphite size.

Special cast iron brake test parts were made with very fine graphite structures and with about 0.04% titanium. Using a hard and abrasive nonasbestos truck brake block, this permanent mold cast iron provided a 30% reduction in the brake lining wear rate, and had a 240% improvement in cast iron life, compared with a conventional brake cast iron.

For the NAO heavy truck brake blocks in particular, the cast iron chemistry and graphite morphology can exert a strong influence on the countersurface wear life. It appears that many nonasbestos materials cause the cast iron to become a more active member of the friction and wear couple. Consequently, closer control of the cast iron may be required. Each new friction material should be tested for cast iron sensitivity, to ensure acceptable service life.

Normal Cast Iron Wear. Usually, a worn cast iron rubbing surface develops a satin gray appearance. Low-abrasiveness passenger car brake linings provide cast iron specific wear rates from about 0.2 to 1 mg/MJ (1.6 to 8 × 10^{-5}in.^{3}/hp·h) in the absence of external contamination or rusting. Countersurface wear and scoring can result from abrasives that are an intended friction material constituent. Some abrasive content to the friction material is desirable, for example, to remove rubbing surface rust after periods of extended nonuse and to control brake lining transfer. Such abrasives also can be used to increase the lining friction level, or to control cast iron crack growth in severe-duty brake linings. Cast iron wear rates from 5 to 16 mg/MJ (4 to 13 × 10^{-4}in.^{3}/hp ·h) can result. At the higher values, the weight loss of cast iron can exceed that of the brake lining wear.

The cast iron wear rate is sensitive to a number of factors. An abrasive will wear the cast iron if it has higher hardness, higher melting-softening temperature, and sufficient particle size. Litharge and barite are both softer than cast iron at room temperature. With normal brake usage, barite will increase the cast iron wear, whereas litharge will not, because barite has a higher softening temperature than cast iron. Even materials that meet the hardness and softening criteria may not produce severe cast iron wear, unless they are of a size that can cause abrasion.

A brake lining matrix hardness varies considerably with formulation and with temperature for a given formulation. Some brake linings soften significantly, and can produce cast iron wear rates as shown in Fig. 3, trace A. Other linings have a nearly constant matrix hardness, and result in wear like that of trace B. The cast iron wear rate increases at elevated temperatures, mirroring the lining wear rate. When abrasives are evenly distributed throughout the brake lining, cast iron

wear rates directly follow the lining wear rates. If external abrasives are present and dominant, wear behavior like that of trace C results. The decrease of cast iron wear at higher usage temperatures then appears to result from softening of the brake lining matrix and abrasive clearance from higher lining wear rates.

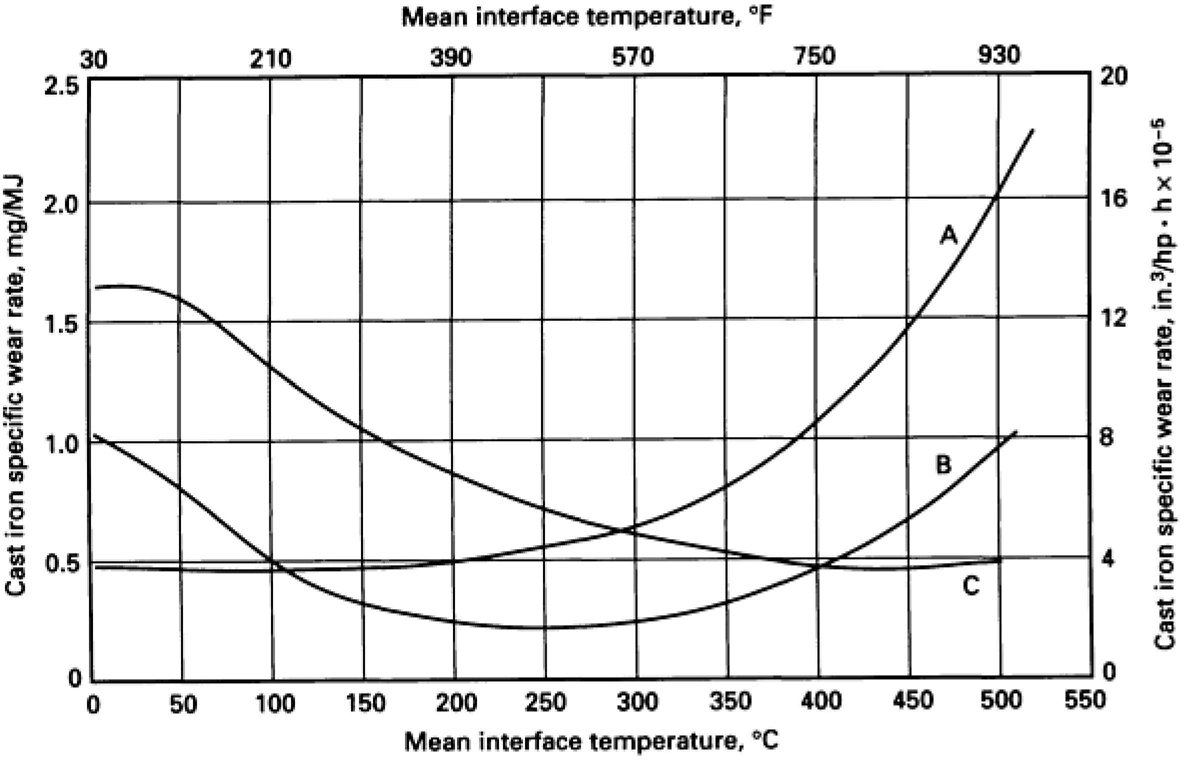


Fig. 3 Cast iron specific wear rates versus brake temperature. See text for discussion.

Local Cast Iron Wear. Brake rotors can generate local brake lining contact zones when there is a significant axial runout. Figure 4 shows the axial runout and thickness variation (TV) of a rotor that had minimal brake usage, but developed a thickness variation from parasitic wear (unintended brake dragging) during highway driving conditions. Road dust contamination aggravated the wear. Similar wear patterns have been seen with semimet brake linings, because of their magnesium oxide abrasive content.

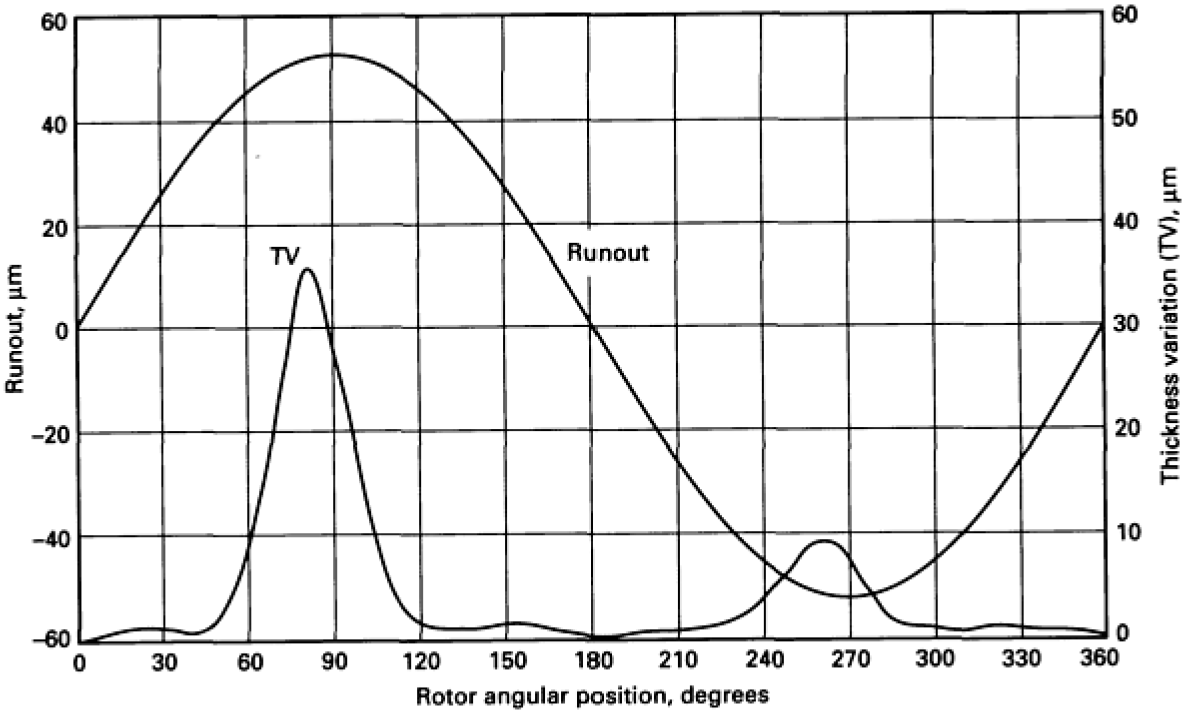


Fig. 4 Rotor axial runout and thickness variation from local drag wear

Highly localized cast iron wear is found on the inboard face of the rotor, centered at a site slightly in advance of the maximum runout position. The resultant brake pedal vibrations from this wear were sufficient to require the remachining of the rotors after only 17,000 km (10,560 miles) of dusty highway service. Often, a small increase of cast iron matrix hardness and hard particle content will reduce the local cast iron wear rate greatly.

Extreme scoring of cast iron can result from an inadvertent contamination of the lining by abrasives. For example, silicon carbide was a "tramp" impurity in the synthetic graphite used in a semimet brake lining formulation. After 2000 km (1240 miles) of customer service, one rotor face was deeply scored over the entire rubbed surface. The opposite rotor face showed no scoring, with the original grinder marks still evident on the surface. Both brake linings were of the same semimet formulation, but from different production batches, one with silicon carbide and the other without. Laboratory testing showed the contamination level at a few tenths of a percent.

Hot spotting of the cast iron can produce local wear at the heated sites. If martensite is formed on the cast iron surface, local high spots will result. Martensite and hot spotting have been addressed by Anderson and Knapp (Ref 9).

External Abrasive Effects. Dust or splash shields can be used to reduce disk lining and rotor cast iron wear from particulate contamination, but attendant restricted brake cooling may increase lining wear rates. External contaminants can be kept from drum brakes by labyrinth seals or shrouding with a small loss of cooling efficiency. However, retention of normal wear products and accumulated rust particles within the brake sometimes then contribute to increased wear rates. With riveted linings, the cast iron wear may be heavily biased along the path of the rivet holes, particularly those at the leading edge of the linings. Careful testing and intelligent compromises may be needed to balance cooling and contamination effects.

Transfer coatings onto the cast iron can be generated by semimet brake linings, as described in the section on brake lining wear. These coatings can vary from a few tenths of a micron to 40 μm (1.6 mils) or more. Once formed, the wear of the original cast iron becomes zero, except from local scoring that is due to abrasive action. However, the transfer coating is by no means static.

With extended operation at low temperatures and especially for rubbing speeds below 2 m/s (6.5 ft/s), the coating will deplete. This appears to result from a preferential transfer to the friction material, because very low lining wear rates are measured during the period when the coating is depleting. Once depleted, the brake lining wear rate increases by a factor

of about 4. The cast iron then wears at a substantial rate, and its appearance changes from a dark blue-black coloration, with a matte finish, to a shiny gray. With time, the cast iron usually develops uneven wear across the rubbing path. During near-transition usage, the cast iron coloration includes bands of brown, blue, and purple.

With continued low-temperature, low-speed usage, the cast iron wear mass removal rate has been found to approach that of the friction material. If these conditions continue, the disk brake linings will wear out at about 20% of the expected mileage, and the rotors will have become unserviceable from the high and generally uneven wear. Normally, such high wear rates are associated with severe usage. In this case, a change of wear mechanism causes high wear under very light-duty, low-temperature usage.

If the cast iron wears under light-duty conditions, but not under hard service, one might redesign the brake to ensure higher temperatures. This measure is acceptable, within limits. Continued high temperature and hard brake usage cause another unique condition. The transfer layer may become so thick that it locally delaminates from the cast iron rotor. This occurs when the transfer layer thickness is about 30 μm (1.2 mils). Figure 5 shows a cracked and incipient failure state for the semimet transfer layer on a medium truck disk brake rotor that had seen severe service on a mountainous test route. Local high spots, around 1 mm (0.04 in.) in diameter, signal these delamination sites. The light bands near the center of the photograph are regions where the coating has broken away locally.

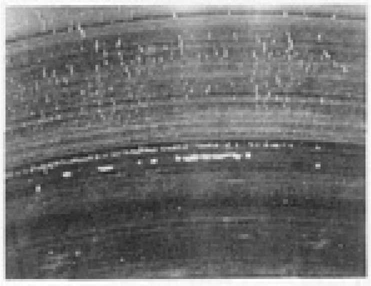


Fig. 5 Semimet transfer layer cracks and blisters

Figure 6 shows a more severe case, with the same semimet lining and brake usage. The cast iron is visible at the bottom of the recesses, where the transfer coating has flaked off. Machining marks could still be seen at these sites. Based on a simple scratch test, the coating was found to be quite adherent. Presumably, the delamination process resulted from high thermal strains at the coating-cast iron interface.

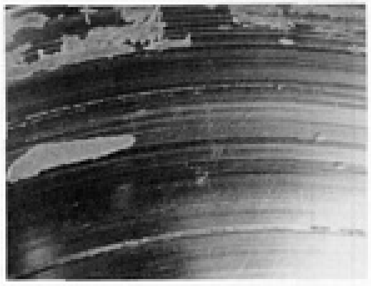


Fig. 6 Semimet transfer layer delamination

Design analysis of brakes for semimet linings thus appears to involve assurance that light-duty brake usage will not produce rotor wear and severe usage will not result in transfer layer delamination. Consequently, optimal design would prevent these usages in most operating situations, and would result in rare occurrences of either wear type. The greater the variation of usage severity in customer service, the greater the possibility of such cast iron wear issues.

Toxicity of Brake Wear Debris

When brake wear debris is considered, asbestos fiber is a common concern. However, several studies of brake lining wear emissions have shown only about 0.03% asbestos in the airborne wear debris. The low emission results from the high frictional contact asperity temperatures. Chrysotile asbestos converts to forsterite at about 800 °C (1470 °F), well below brake flash temperatures. Organic brake lining fiber emissions averaged 0.7 μm (28 μin.) in length. Therefore, the vast majority of the airborne asbestos did not meet the definition of a fiber by the Occupational Safety and Health Administration, the World Health Organization, and other authorities. Epidemiological data for automotive brake mechanics have shown no detectable excess cancer risk. The Environmental Protection Agency, using computer models for fiber exposure and cancer risk, predicted about 15 excess cancer cases a year from all U.S. brake wear.

Semimet brake wear emissions appear to be of minimal toxicity. Fibers from the wear of other nonasbestos brake linings have not yet been collected and measured. These include mineral wool, glass fiber, slag wool, titanate fiber, phosphate fiber, carbon, and para-aramid fiber.

Although lead was removed from most brake linings by 1980, barium, usually in the form of barium sulfate (barite) powder, is often found. Such heavy metal wear particles may affect chest x-rays of brake mechanics. Brake usage generates some fused-ring aromatic material, but in miniscule quantities, like those in lettuce. Until complete and definitive brake wear toxicity studies are completed, it would be prudent to minimize dust exposure from any brake wear debris.

Automotive Brake Frictional Characteristics

Brake linings inherently have some performance attributes, such as fade and fade recovery, but full evaluation of a friction material requires that it be installed into a full brake system. Vehicle brakes are required to operate under a wide range of conditions, from hard braking with a heavily loaded vehicle on a steep downhill slope to minimal brake usage on interstate highways. Vehicle brakes should be highly reliable and minimally affected by temperature, water, or other contaminants. Brake lining friction must be consistent throughout the life of the material.

Brake Design Basis. Brakes are designed primarily on the basis of wear, pedal travel, brake system stability, and effectiveness properties. Brake effectiveness is defined as the ratio of the brake friction torque to the applied force. This term is used, rather than friction coefficient, because brake geometry factors often make the brake torque not linearly related to friction level. For example, effectiveness must be used to describe frictional performance of drum brakes, because of their large self-actuation characteristics.

Self-Actuation. Frictional forces acting on the shoes of a simple drum brake may cause it to be further loaded against the drum, increasing its effectiveness. Such shoes are referred to as leading. Friction force on a trailing shoe causes it to oppose the application force, decreasing effectiveness. The increase in brake shoe loading, which is due to friction and geometry effects, is referred to as self-actuation.

Although the geometric considerations of self-actuation influence brake effectiveness, it also is influenced by frictional properties of the brake lining. Figure 7 shows design effectiveness curves for different brakes as a function of friction coefficient. Assuming a nominal friction coefficient of 0.4 for each 1% change in friction coefficient, the brake effectiveness changes by approximately 3.5% with a duo-servo drum brake (two leading shoes, coupled), and by 2.6% on a leading-trailing drum brake, but the change is just 1% for a nonservo disk brake.

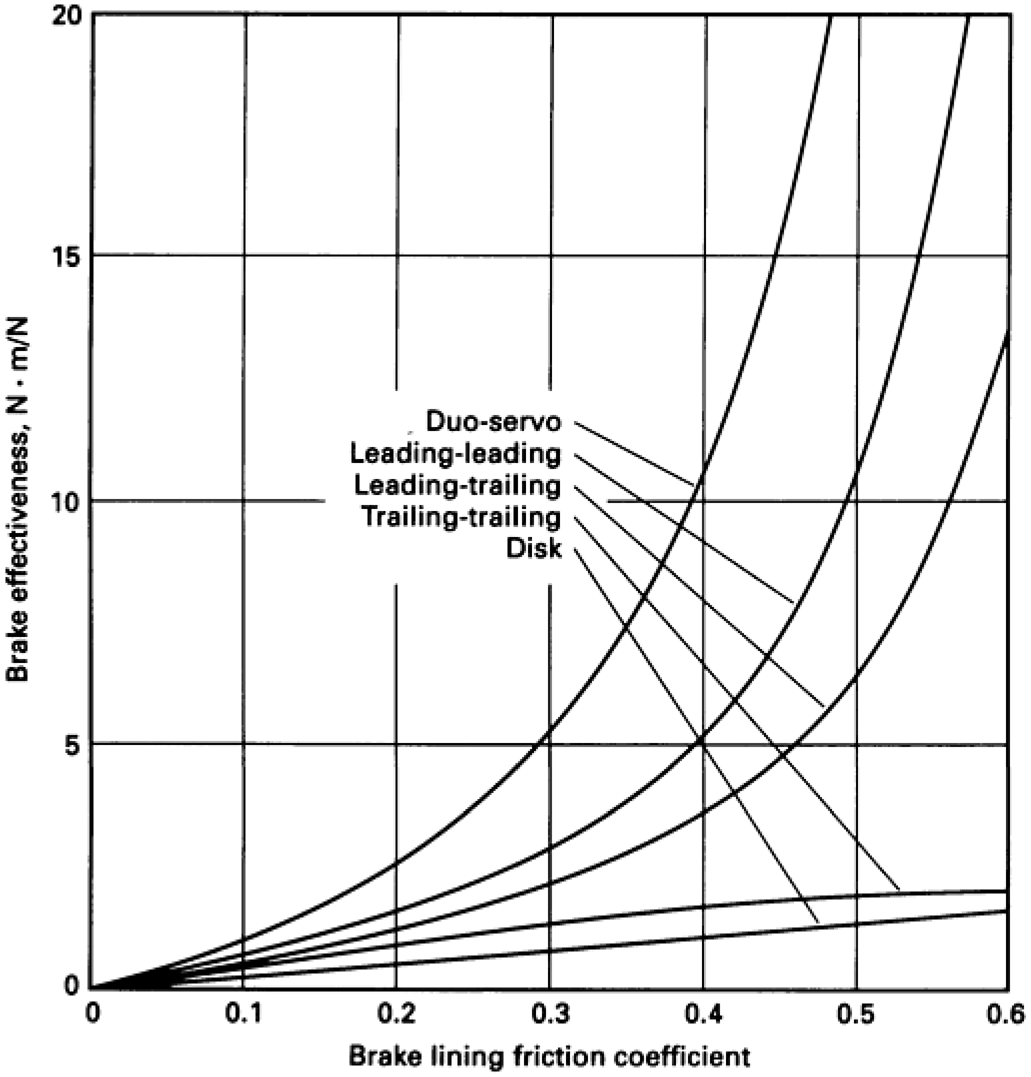


Fig. 7 Effectiveness versus friction coefficient for several brake types

Design Factors. In addition, the design of brakes can affect the degree to which friction material wear and thermomechanical properties influence brake effectiveness. A large change of brake effectiveness is possible, for the same value of friction coefficient, simply because of a change of lining pressure distribution. Figure 8 shows the change of effectiveness on a leading shoe that results from different pressure distributions. For a given value of friction coefficient, the effectiveness may vary by a factor of 4 as the pressure distribution varies from center-biased to end-biased. On a drum brake, the lining pressure distribution is affected by the stiffnesses of drum, shoes, and brake linings, in addition to thermal distortions, wear patterns, and prior usage history effects.

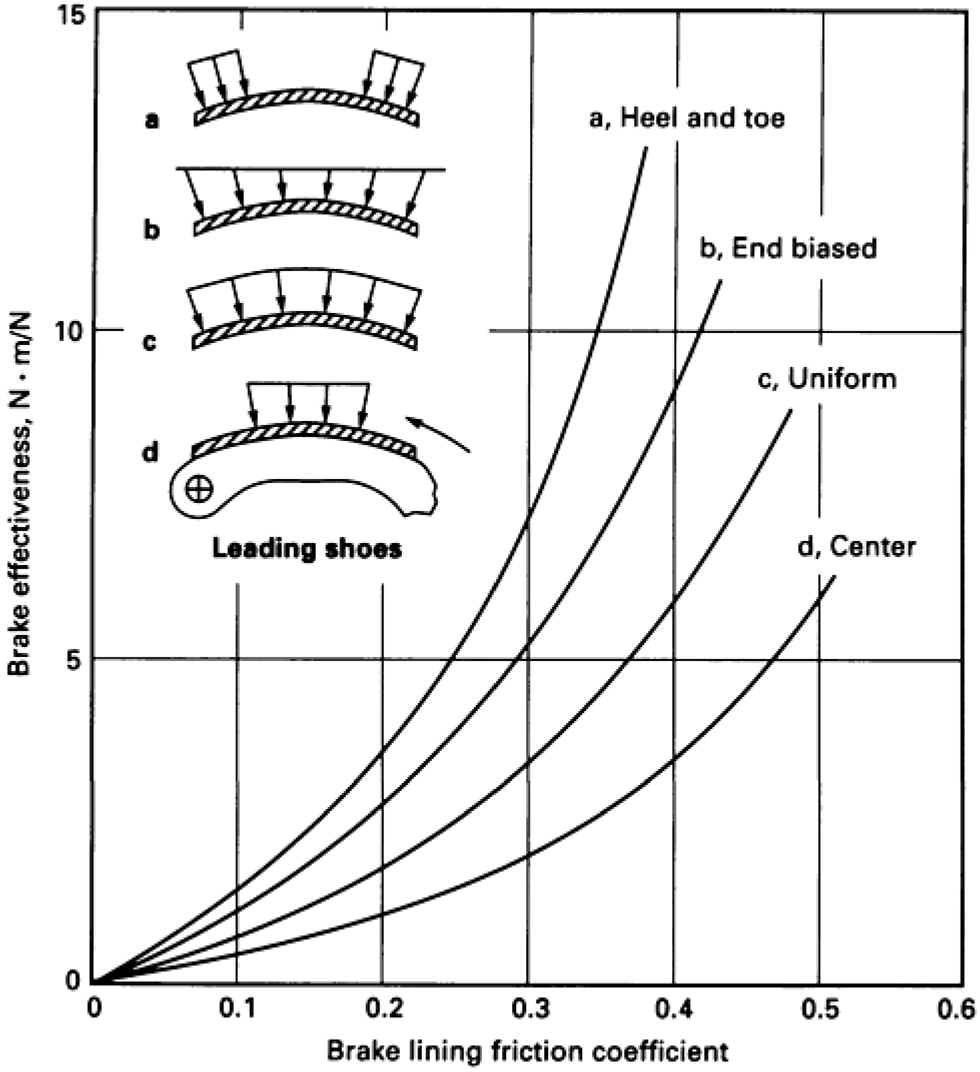


Fig. 8 Effectiveness variation with pressure distribution, based on type of leading shoe

The brake designer can achieve the same brake effectiveness by several combinations of lining geometry, brake shoe geometry, and lining coefficient of friction. Thus, changing friction material type (as from asbestos-organic to one of the nonasbestos types) may provide significant effects on brake system performance, because of different coefficients of friction, different thermal expansion coefficients, and different compression stiffnesses.

Many brake design parameters are interrelated. For example, a large value for lining thermal expansion may require added running clearances to avoid parasitic drag. This results in greater brake actuator travel, which may then necessitate further brake system redesign. A complete introduction to brake design is beyond the scope of this article. The preceding discussion was presented to introduce design issues and to illustrate the virtual impossibility of simple substitutions.

Brake Frictional Performance

Some friction material effectiveness measures that are evaluated during new vehicle brake system qualification experiments include: burnish and green performance, fade resistance, fade recovery, delayed fade, speed sensitivity, effectiveness drift, and environmental sensitivity. Each of these important performance characteristics is further described below. These are the most important of a much larger set of brake performance attributes that are evaluated during brake system development.

Green Effectiveness. When a friction material is new, it is unburnished and its contact with the drum or disk depends on many manufacturing tolerance stack-ups. With a new brake assembly, effectiveness often is lower and more variable than after the friction material is fully burnished. This initial behavior is called the green effectiveness.

The initial brake lining friction values may vary because of surface texture effects, surface contamination effects, contact pressures, and brake lining composition. Particularly with drum brakes, the green effectiveness also depends on the initial contact geometry. Design dimensions are chosen to provide lower initial effectiveness from the contact geometry. This minimizes the possibility of the overeffectiveness of new brakes, which is due to either rust or some other contaminant that increases initial friction.

Burnished Effectiveness. After the drum or disk has contact over the full rubbing surface, and when a stable char layer has formed over the entire brake lining contact area, the brake is said to have been fully burnished. Typically, the brake performance has stabilized by the time that 85 or 90% of the brake lining has contacted the brake drum or disk. An exception to this exists for the semimet brake lining. Semimet frictional behavior is discussed separately, because it has several unique characteristics.

Fade Resistance. Brake fade refers to a loss of brake effectiveness, generally as the result of excessive brake temperatures. However, five types of brake fade can occur:

- Thermal, which is due to high brake bulk temperatures
- Delayed, which is due to resin migration during cooling
- Blister, which is due to near-surface lining blisters
- Flash, which is due to high-speed, high-load braking
- Contamination, which is due to partial lubrication from water or oil

Poor brake fade behavior could be exhibited by high-quality materials, if used is inappropriate brake applications. Brake lining fade behavior is meaningful primarily in the context of a particular brake lining, brake, and vehicle usage situation. Depending on the brake usage sequence and lining fade response behavior, a driver may either sense impending fade as a gradual increase of brake pedal force requirement or he may not have any warning. Good brake lining choices provide a gradual and progressive loss of effectiveness to warn of impending brake fade. Good applications do not provide the driver with surprises.

Fade recovery refers to the ability of the friction material to quickly regain normal brake effectiveness after thermal fade. As the brake disk or drum cools, the friction level should return quickly to the prefade level.

Delayed fade is a phenomenon that can occur with some drum brake friction materials. During the fade recovery, brake effectiveness may drop unexpectedly, causing a temporary, but pronounced, increase of brake pedal force requirement. A delayed fade is insidious in that it is unexpected. It occurs well after a period of hand brake usage and usually with no warning signs.

Blister Fade. New brake linings may contain volatile material from fabrication, which, if not released by the end of the burnish process, could cause high internal gas pressures upon rapid heating, as during a hard brake application. In some situations, a near-surface blistering results in a rapid, brief loss of brake effectiveness. Friction is lost because of excessive contact pressures at the blister sites and from the evolved gases. A characteristic of blister fade is brevity. Effectiveness is lost for a few seconds during a hard brake application, and then returns to near-normal. Repetition of the hard brake application will not produce a second blister fade, because the volatile material has been eliminated from the near-surface region.

Flash fade is related to the blister and green fade, but occurs only at very high brake power levels, usually at very high speeds. The rapid decomposition of near-surface organic constituents produces a gas-pressure-lubricated braking surface. The brake lining friction may not be low, but the evolved gas pressurizes the friction material to counteract the applied force. High surface area brakes, and those with high organic contents, are most vulnerable. Prior brake usage at moderate-to-high brake lining soak temperatures reduces flash fade severity.

Contamination Fade. Water, oil, or a combination of these on the surface of the brake lining or brake drum/disk can generate an elastohydrodynamic fluid film that effectively makes a bearing from a brake. Different friction materials have

different porosity, compliance, and wear characteristics, and thus may be quite different in sensitivity to contamination. High surface area brakes, such as drum brakes, are more prone to contamination fade. However, even automotive disk brakes can exhibit such a fade, if saturated by either water or oil/water. This effect is similar to that from oil/water films on the road surface, affecting tire friction.

Speed Sensitivity. All brake linings exhibit different frictional behavior as the rubbing speed changes. The speed sensitivity varies with material type and composition. It also varies with temperature, pressure, and prior usage history. To ensure proper brake balance over a wide range of stopping speeds, the brakes on each axle should exhibit similar effectiveness characteristics.

In general, brake effectiveness decreases with increasing speed. Disk brakes provide less speed spread, or difference in brake effectiveness at different braking speeds, than do drum brakes. This is primarily due to the low servo factor of disk brakes. Duo-servo drum brakes, with high self-energization, are the most sensitive to speed. Good front-to-rear brake balance at all speeds is difficult to attain on vehicles with front disk brakes and duo-servo rear drum brakes. Highly speed sensitive disk brake lining need to be paired with low-speed-spread drum brake linings to compensate fully for the 3- to 4-fold difference in brake servo factors.

Environmental Sensitivity. Contaminants such as water, oil, dust, and rust alter brake effectiveness. For example, contaminant oil films of 0.2 μm (8 μin.) thickness can raise the brake friction coefficient by 55% at low vehicle speeds. This can make some drum brakes virtually self-locking. However, at 0.5 μm (20 μin.), an oil film can lower friction by elastohydrodynamic action. Because of low oil tolerance and high sensitivity, even an oily fingerprint can cause significant effectiveness change.

Wet Friction. The performance of vehicle brakes when wet also represents a safety concern. Disk brakes usually are less affected by water than are drum brakes, largely because of their lower inherent servo factor. However, both disk and drum brakes may show large effectiveness losses when wet. A small amount of oil, together with water (as with tires) increases the lubricating action.

When wet, some friction materials provide a greater loss of brake effectiveness than others. These usually also take a considerably longer time to recover friction capability after wetting. Permeability, homogeneity, and compression stiffness are some of the brake lining properties that help determine wet friction response. A complete understanding of this behavior is not known. Full brake dynamometer and vehicle tests are used to establish the wet friction behavior of brake systems.

Moisture Sensitivity. Friction materials are typically somewhat porous, fiber-reinforced composites that are capable of absorbing atmospheric moisture when a vehicle is parked, such as overnight. This moisture has been shown to change break effectiveness temporarily for some friction materials. Glass and para-aramid ten to increase friction when moist. Other materials tend to lower friction in moist environments.

Drum brake linings also change in thickness with time when exposed to high-humidity conditions. When the brake lining ends thicken, relative to the center, brake effectiveness increases.

Rust Effects. Another moisture effect results from rusting to the cast iron disk/drum brake lining surface when a vehicle is parked for some time, causing "stiction." Under adverse environmental conditions, it is possible for some friction material to rust-bond to the cast iron. A substantial torque can be required to break this bond. Thereafter, the rusted surface of the brake drum or disk may generate an uneven brake torque with brake angular position.

Rusting of the brake drum or rotor can occur independent from rust-bonding. In this situation, the brake effectiveness tends to be high at low speeds and low-to-normal at higher vehicle speeds. Whether rust prevails or not depends on the relative abrasiveness of the brake lining and rotor surface, as well as the brake usage conditions.

Although some of the above characteristics can be investigated in the laboratory, environmental testing is required to establish acceptability of friction material formulations in full vehicle service usage. The ideal brake assembly exhibits little or no moisture sensitivity. When present, it usually persists for a few brake applications, and then disappears when brake heat drives the moisture from the brake lining or when wear removes the surface rust. A small abrasive content in a brake lining may be beneficial in removing minor rusting from the brake drum or disk. Brake lining formulations also can contain agents to control minor rusting in service.

Effectiveness Drift. To ensure consistent vehicle braking performance, the brake effectiveness characteristics should be stable throughout the life of the brake linings. The brake effectiveness of some friction materials can deteriorate with accumulated usage history, time, and wear. The formation of a surface glaze on some brake linings, during prolonged light-duty usage, has been associated with a loss of effectiveness. Linings with low wear rates tend to glaze more readily that do those with poor wear life.

Semimet Frictional Behavior

Resin-bonded metallic, semimet brake linings have sufficiently different performance characteristics that it is simpler to describe them separately. Disk brake effectiveness with semimet brake linings is virtually constant with normal usage pressures and temperatures, for speeds ranging from 50 to over 200 km/h (30 to over 125 mph). Effectiveness drops slightly below 50 km/h (30 mph).

Thermal Fade. Semimet linings tend to be nearly fade-free. When thermal fade is experienced, it occurs as a progressive loss of effectiveness with decreasing speed. Significant brake fade primarily occurs at the end of the stop.

Environmental Sensitivity. Semimet linings are prone to have a 50 to 60% loss of effectiveness after exposure to moist air for one hour or more. This becomes the limiting condition and determines the amount of power brake boost required with most semimet brake linings (most other brake lining types are limited by initial fade behavior). It also restricts the available choices for compatible drum brake linings.

Water Sensitivity. Water affects many friction materials greatly. Semimets are minimally affected unless the water is accompanied by an oil film with a thickness of 1 μm (40 μin.) or more. In that case, the frictional effect is similar to other brake linings. However, disk brakes run warm enough to dry quickly, so water effects usually are neither critical nor long lasting.

Laboratory and Vehicle Brake Evaluation

Laboratory specimen testing machines are commonly used to characterize and audit the quality, or sameness, of friction materials, using specimens from full brake linings. The advantages of using laboratory systems for evaluation include automated testing, careful control of operating conditions, and more rapid measurement of brake lining characteristics. In addition, laboratory evaluations that utilize a specimen test are less costly than full-scale brake dynamometer or vehicle experiments.

Several different laboratory machines are used to determine brake lining properties. Three commonly used laboratory test machines are described below.

The friction assessment screening test (FAST) machine was developed for rapid "fingerprinting" of friction material specimens. The FAST machine is used primarily for routine brake lining quality control testing. Although this machine reportedly has been used to perform friction material screening tests, it never has been recommended as a substitute for full-scale brake evaluations.

A friction materials test machine is also used for quality-control testing. The Society of Automotive Engineers (SAE) has designated a recommended practice (SAE J661a), used in conjunction with this machine, that is classified as a quality-control test procedure. It is used more for periodic quality-control surveillance testing, as opposed to routine production batch testing. Based on test data from the SAE J661a procedure, a brake lining friction rating specification (SAE J866a) has been used by some states to classify and regulate friction materials. However, this rating system has been shown to be clearly inadequate for meaningful comparative testing with different types or classes of friction materials. SAE J866a (revised in 1984) includes a caution against such uses.

A full brake inertia dynamometer simulates vehicle braking by mounting a complete brake assembly to a large rotating inertia and drive motor. Most are single-ended, and test a single brake assembly at a time. Brake dynamometers place the brake assemblies in closed ducts, both to expedite cooling and to control smoke and smell. Because faster cooling rates hasten testing, most dynamometer tests have much greater air flow, and resultant cooling rates, than are found in on-road testing. Consequently, brake dynamometers are used mostly for controlled wear tests, basic effectiveness tests, initial fade/recovery tests, and parking brake tests.

Full brake dynamometers are available in a range of sizes, with inertial equivalents from motorcycles to railroad locomotives. Because there are no brake dynamometer standards, many are unique. Dynamometer-to-dynamometer differences in test results can be significant, even for carefully matched linings and brakes that are tested to the same inertia loading. As a result, most brake test engineers prefer comparative brake lining test data, obtained from a single brake dynamometer, to data from different brake dynamometers.

Full brake dynamometers can closely replicate most vehicle in-service braking conditions, except for air flow over the brake and environmental conditions, such as water and dust contamination. They are therefore invaluable for brake diagnostic testing, and can be excellent for initial brake lining screening tests. However, they are not sufficient to fully evaluate brakes.

Correlation of Laboratory and Vehicle Test Results. Numerous studies have been done to determine the degree of correlation between laboratory friction material test results and actual vehicle test results. In general, the only good analog of a vehicle brake is the brake itself. Consequently, there are no specimen or scale test devices that consistently yield test data that correlate with full vehicle data. This does not mean that such laboratory tests are useless, just that they should not be used to predict field performance behavior. They can, and have, been used to screen friction materials with performance flaws, but require expertise in data analysis.

Even full brake dynamometers are difficult to correlate with on-road vehicle brake performance, unless careful instrumentation and test controls are used. It should be possible to correlate full brake dynamometer test results with on-road data without a need for questionable correction factors. Although this has been tried by many, and published by a few, it has not as yet been verified by anyone.

The testing of brakes involves evaluations under the widest possible sets of conditions, in order to find flaws and behavior inconsistencies. Friction material formulation and testing both are more of an art than a science. If accurate models of brake performance were available, then testing would be much simpler. However, each friction material formulation is proprietary, and can be unique in many respects to other materials. Consequently, brake test engineers use a combination of standard performance test procedures to ensure adequate capacity. Customer service fleet tests over the entire United States then help to determine usage and environmental stability characteristics.

Prudent brake development engineers start with friction materials that have been fully screened by laboratory tests, and then run full brake dynamometer tests to further screen the candidate brake linings and to initiate brake hardware adjustments. Only then are vehicle performance tests justified. When these are acceptable, the more difficult, costly, and time-consuming fleet and traffic tests are initiated.

Federal Braking Requirements and Other Brake Tests. As indicated, the qualification of vehicle brake systems and friction materials can involve numerous experiments to determine fade resistance, moisture sensitivity, wet friction, and other performance parameters. On new vehicles, brake performance must meet Federal Motor Vehicle Safety Standards (FMVSS) 105-75 for hydraulic brakes and 121 for air brakes. Other performance standards are determined by the vehicle and brake manufacturers. Replacement brake linings have no legislated performance standards to meet.

SAE-Recommended Practices. The SAE has developed about 20 recommended practices for checking the performance of brake linings and systems. These standards cover automobile, truck, and trailer brake system tests using both vehicles and dynamometers.

Prior to the adoption of FMVSS 105 and 121, the SAE procedures were intended to provide guidelines to brake system evaluations. The passage of FMVSS 105 as a requirement for brake system certification shifted the emphasis of these SAE procedures to the role of supplementary tests that could be used to further qualify vehicle brakes and braking systems. The majority of the SAE brake codes have been directed toward automobiles and light trucks, although some (SAE J880 and SAE J9781) have been developed for heavy commercial vehicles.

Friction material performance in vehicle brake systems can be qualified by meeting the requirements of the federal safety standards and being evaluated by the SAE recommended practices, along with brake component and vehicle manufacturer standards. For new vehicles, compliance with federal and state motor vehicle safety standards is only a starting point for the acceptance of a brake system. No federal standards exist today for replacement brake linings.

References

1. S.K. Rhee and P.A. Thesier, "Effects of Surface Roughness of Brake Drums on Coefficient of Friction and Lining Wear," Preprint 720449, 1972222, Society of Automotive Engineers, 1937
2. H. Blok, General Discussion, *Proc. Inst. Mech. Eng.,* Vol 2, 1937, p 222
3. S.C. Lim and M.F. Ashby, Wear-Mechanism Maps, *Acta Metall.,* Vol 35, (No. 1), 1987, p 1-24
4. S.K. Rhee, "Influence of Rotor Metallurgy on the Wear of Friction Material in Automobile Brakes," Preprint 710247, Society of Automotive Engineers
5. M.H. Weintraub and J.P. Bernard, Chemical and Functional Responses to Brake Lining Cure Variations, *SAE Trans.,* Vol 77, paper 680416, 1968
6. M.H. Weintraub, A.E. Anderson, and R.L. Gealer, Wear of Resin Asbestos Friction Materials, *Adv. in Polymer Friction and Wear,* Plenum Press, 1974
7. A.E. Anderson, Wear of Brake Materials, *ASME Wear Control Handbook,* American Society of Mechanical Engineers, 1980
8. D. Hatch, Cast Iron Brake Discs, *J. Automot. Eng.,* Oct 1972, p 39
9. A.E. Anderson and R.A. Knapp, "Hot Spotting in Automotive Friction Systems," presented at 1989 International Wear of Materials Conference and published in *Wear,* Vol 135, 1990, p 319-337

Friction and Wear of Tires

P.S. Pillai, Goodyear Tire & Rubber Company

Introduction

TIRE WEAR is a complex phenomenon that occurs in the tire footprint area--that is, the interface region between tire tread and pavement surface. The forces required to support and control an automobile act in the footprint contact area. Forces generated in the tire footprint allow the vehicle to change direction and speed of travel. Production of these action and reaction forces generates frictional work in the footprint region. The dissipation of this frictional work is the primary reason for tire wear. Factors that affect tire wear are broadly classified in Table 1.

Table 1 Parameters that affect tire wear

Tire-related factors	Operating conditions	Other factors
Tire type	Tire load	Pavement texture and condition
Tread design	Inflation pressure	Seasonal effects
Tread composition	Speed	Ambient temperature
Tire construction features	Driver vehicle interaction	
Magnitude of inertial forces	Vehicle maneuvering behavior	

Abrasion and Wear

Tire wear or tread wear refers to loss of tread rubber. Abrasion of rubber refers to the removal of rubber particles from a testpiece (a small rubber wheel, for example) when rubbed against another hard object. Thus, it would seem logical to assume that tread wear is related to rubber abrasion. However, experiments have shown that the two processes do not exhibit a one-to-one correlation. Informative review articles on rubber abrasion have been published (Ref 1, 2, 3, 4, 5).

Basic research on rubber abrasion and wear has been performed by Schallamach, Gough, Grosch, and other investigators (Ref 6, 7, 8, 9, 10, 11). Study of rubber abrasion is essentially limited to laboratory testing of model rubber specimens using abrasion machines; wear refers to tires in service. Many authors have tried to use laboratory abrasion results to predict tread wear, but have had only limited success. In the tire wear process, although te physical phenomenon is loss or abrasion of tread rubber, many factors--tire type, design and construction parameters, material properties, vehicle maneuvering behavior, driver-vehicle interaction, weather, and so on--influence wear rate. Therefore, direct wear measurements are required to rank the wear rates of different tires. The tread wear grade provided by tire manufacturers is a comparative rating based on the wear rate of the tire when tested under controlled conditions on a specified government course. Veith and co-workers have performed extensive experimental work on tire wear (Ref 12, 13, 14, 15). Veith has also published a comprehensive review of the topic (Ref 16).

Definition of Terms. Important terms related to tire wear are defined below.

- *Tread loss:* Average cumulative loss in tread depth for all the grooves in the tire, expressed in millimeters
- *Rate of wear:* Tread loss in millimeters per 10,000 km of travel
- *Tread wear index:* (Wear rate of reference or control tire)/(Wear rate of experimental tire) × 100
- *Standard wear rate:* mils per 1000 miles per 100 lbf cornering force at tire test temperature
- *Camber angle:* The vertical angle between the wheel plane and a line perpendicular to the pavement (Fig. 1)
- *Slip angle:* The horizontal angle between the tire circumferential midpoint plane and the direction of tire motion (Fig. 1)
- *Cornering stiffness:* Force per degree of slip angle
- *Test severity:* Function of tire force severity, pavement surface severity, or weather/temperature severity

- *Tire force severity:* Magnitude of cornering force
- *Pavement surface severity:* Micro- and macrotexture, blunt or harsh (qualitative ranking)
- *Weather severity:* Summer/winter temperature, rain, snow
- *Severity index gradient:* Rate of change of tread wear index with tire force severity
- *Driving severity number (DSN):* Sum of the squares of lateral acceleration per tire revolution (g_i) divided by the number of revolutions (N) normalized for vehicle weight:

$$DSN = \sum(g_i/100)^2\,(F_z F_{z,R})^2/N \qquad \textbf{(Eq 1)}$$

where F_z is the actual tire load and $F_{z,R}$ is the rated tire load.

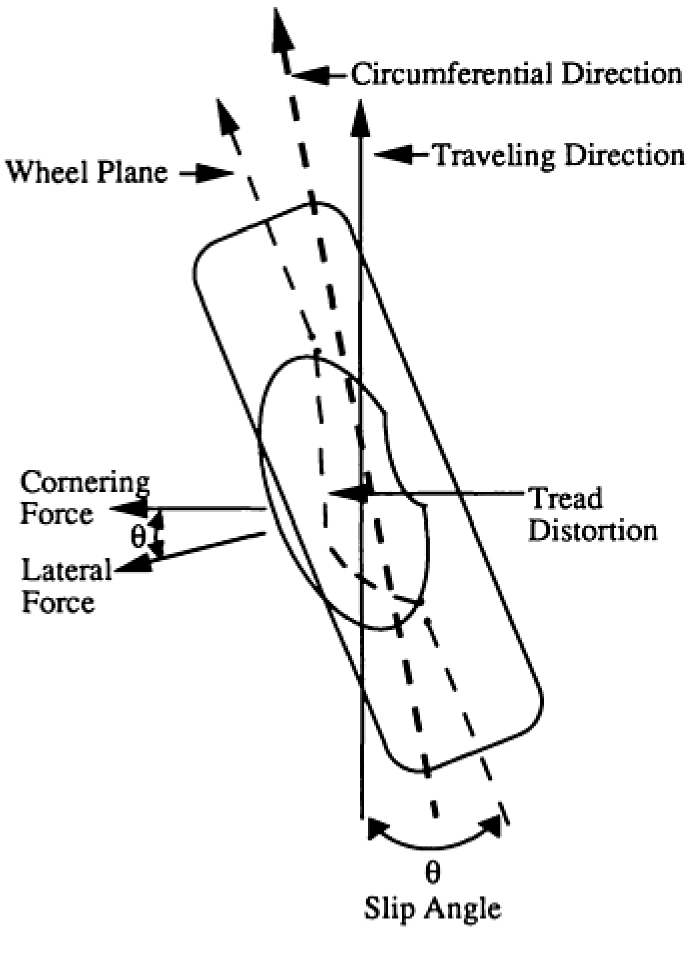


Fig. 1 Schematic showing tire footprint forces and tread distortion

Basic Mechanism. Figure 1 is a schematic of a tire making turn, the direction of motion, the slip angle, and the cornering and lateral forces are indicated. The tire is skewed to the left by the slip angle θ. The distortion of the tire footprint generates a lateral force, mv^2/R, so that the vehicle can make the turn (m is the mass on the tire, v is the instantaneous velocity, and R is the radius of curvature). The tread rubber is held to the road by rubber friction at the leading edge of the footprint. The forces acting in the footprint distort the tread area laterally. The distortion increases in magnitude as the tread element travels from the leading edge toward the trailing edge of the lateral footprint. When the local lateral distortion exceeds a certain limiting value, the frictional force cannot maintain the deflection and the tread element slides back toward the tire center plane. The tread rubber stores elastic energy while in the distorted state. This stored energy is partially converted into frictional work during the sliding domain. The dissipation of this frictional work during the cornering mode is the source of free-rolling, untorqued tire wear.

Tire Axis System. Figure 2 includes additional parameters (for example, tire forces, tire moments, and tire angles) that affect tire wear.

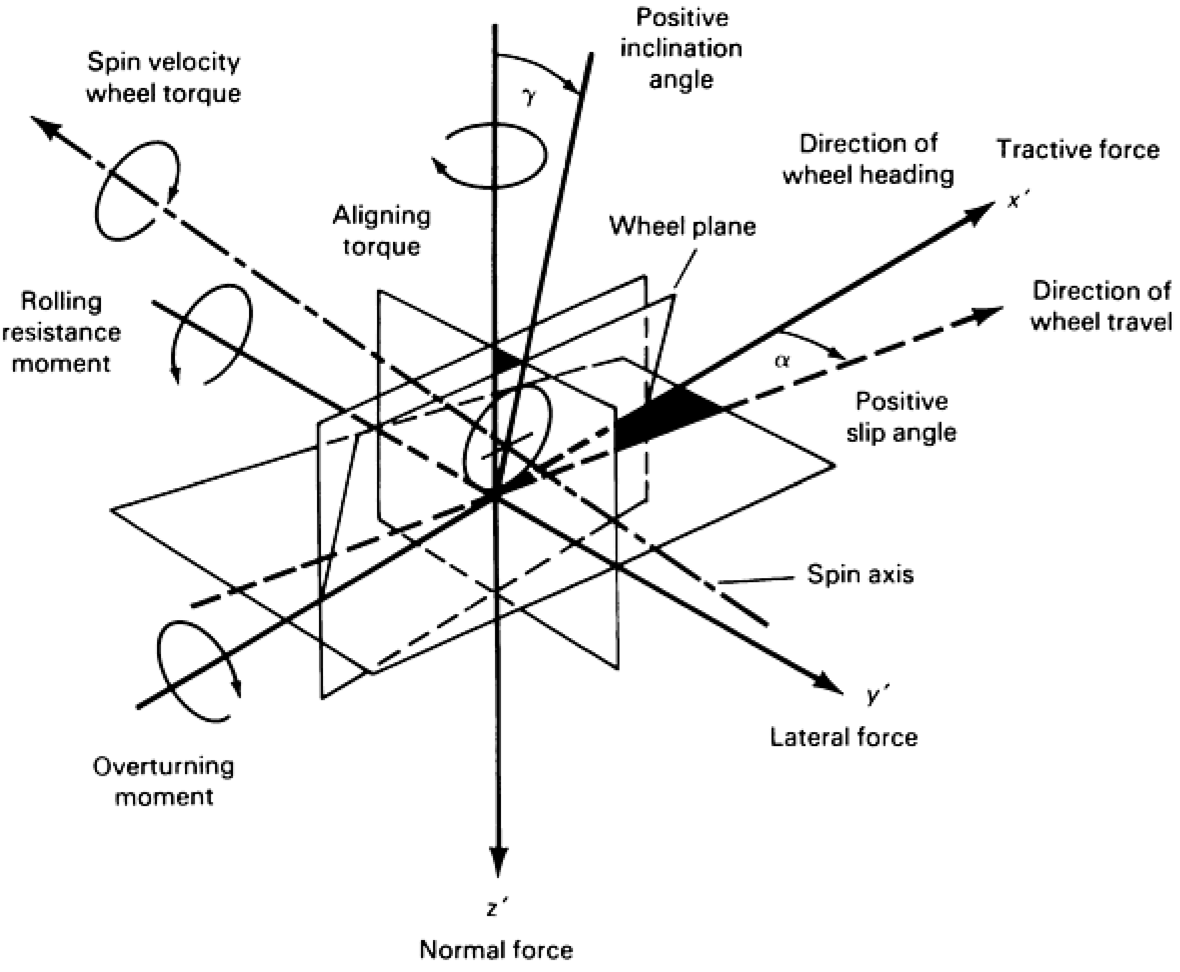


Fig. 2 Schematic showing angles, moments, and forces which comprise a tire axis system. Source: Ref 17

Tire Wear Models

Theoretical Model. Using basic principles of physics, Schallamach and Turner (Ref 18) developed an equation for even wear of tire treads in terms of average cornering force (*F*), cornering stiffness (*C*), tire resilience (*R*), and tread rubber abradability (γ):

$$A = \frac{\gamma R F^2}{C} \cdots \qquad \textbf{(Eq 2)}$$

where *A* is the wear rate per unit distance. Pulford (Ref 4) discusses some of the practical implications of Eq 2. Livingston (Ref 19) extended the Schallamach-Turner model by partitioning the total energy stored into two parts--a fraction stored in the tire body as a whole and the remainder stored in the tread--and obtained the following expression for tread life:

$$h = \left(h_0 + \frac{B}{A}\right)\exp(-Ax) - \frac{B}{A} \qquad \textbf{(Eq 3)}$$

where h and h_0 are final and initial tread height, respectively, x is distance traveled, and B and A are experimental constants.

Empirical Model. Experimental results have shown that the theoretical model expressed in Eq 2 must be modified as:

$$W = kF^n \qquad \textbf{(Eq 4)}$$

where k is a constant and the exponent *n* ranges from 2 to 4. The value of *n* depends on the slip angle, the nature of the pavement, and carbon black particles (Ref 12). The Schallamach-Turner equation (Eq 2) indicates that force is the controlling factor in tire wear. Gough (Ref 11) has also shown that cornering force is the major source of tire wear. At equivalent force levels, cornering action causes higher energy loss in tires than braking action. Experimentally, it has been shown (Ref 12) that tire wear varies exponentially, rather than linearly, with force. This is general agreement with Eq 2 and 4.

The above models explain regular wear in tires reasonably well. However, tire also exhibit irregular wear, and at present there are no good mathematical models to explain this phenomenon. According to Livingston (Ref 20), irregular wear may occur because of differences in stiffness and ground pressure from point to point in the tire footprint area. These differences result in different strain levels within the contact area, leading to different amounts of sliding from point to point. Livingston has proposed a theoretical approach that assumes differential slips for various tread elements to explain irregular wear. He suggests using finite element analysis (FEA) simulation in incremental steps.

Standard Wear Rate

Tire design and construction parameters affect the nature of the forces generated in the footprint area. As mentioned earlier, force is the dominant controlling factor in tire wear. Therefore, to compare wear rates for different tire constructions, the measured wear rate for each construction can be related to a reference force value and corrected for surface temperature variation. This corrected rate is known as the standard wear rate. Incorporating Schallamach's temperature-correction relationship (Ref 8), Veith (Ref 12) developed the following equation for the standard wear rate:

$$W^{\circ}_{50} = \frac{[W \cdot (100)^2]}{F^2[1 + \alpha(T - 50)]} \qquad \textbf{(Eq 5)}$$

where W°_{50} is the standard wear rate at 50 °C (120 °F) and 100 lb cornering force, *F* is the average cornering force in pounds, α is the temperature coefficient of wear rate, and *T* is the test temperature in degrees Celsius.

Fleet Testing and Trailer Testing

Measurement of tire wear is generally done by fleet testing. A car fitted with one type of tire is driven over specific routes representative of typical driving conditions. Test conditions, such as tire rotation, frequency of vibration, speed, straight driving, turns, and nature of pavement, are preset. The initial and final tread depths averaged over all the grooves are measured, and the total wear is calculated from the difference between the tread depth values. Based on the miles driven, an "average" wear rate is obtained. The test is repeated using tires with different variations in design, compound, materials, and so on. Results allow comparison of the wear rates of different tires. The primary limitation of this method is that the various factors that influence tread wear cannot be independently controlled.

In the trailer testing method, both cornering force and slip angle can be controlled and monitored. This method allows the researcher to directly control the magnitude of the force acting on the tire. Thus, wear rate as a function of force can be determined. Veith (Ref 12) describes a trailer test system with elaborate instrumentation for controlling test parameters and for making measurements. He used this method to study the effects of tire force, weather conditions, and pavement texture on wear rate. He also showed that wear rate is proportional to F^2, where *F* is force, which is in agreement with Eq 2. Driving severity number (*DSN*) is a concept developed by Veith (Ref 21) to arrive at a composite parameter value that combines the total effect of the three factors mentioned above. Using the *DSN*, Veith (Ref 21) quantified the tire input force and driver vehicle interaction over a test course to better explain tread wear rankings.

Effects of Load and Velocity

Veith (Ref 12) has shown that lateral force, *F*, can be expressed in terms of vehicle weight, *L*, as:

$$F = \left(\frac{L}{G}\right)\left(\frac{V^2}{R}\right) \quad \textbf{(Eq 6)}$$

where V is vehicle velocity, R is the radius of curvature, and G is acceleration due to gravity. L is equal to four times the average tire load. Combining Eq 6 and Eq 2, the wear rate can be written as:

$$W = K\left(\frac{L^2V^4}{R^2G^2}\right) \quad \textbf{(Eq 7)}$$

where K is a constant. This indicates that the wear rate is proportional to the second power of tire load and to the fourth power of vehicle velocity.

Experimental verification of Eq 7 is not straightforward. It has been found that the extent of change in W with change in L depends on how W is measured--by trailer method or by vehicle cornering. For the same change in ΔL, the corresponding change in cornering force is much smaller in the former test procedure. Therefore, change in W is higher in the vehicle cornering method than in the trailer method.

Again, the relationship between wear rate and vehicle velocity also seems to be method dependent. Veith (Ref 12) found that the effect of vehicle speed was very small in trailer testing and had a more pronounced effect in fleet testing. Experimentally, it was found that the wear rate, W, was approximately proportional to the cube of the velocity (Ref 4). It appears that there is some inconsistency between theory and experiment, as well as between different experiments, and it is thus difficult to correlate experiment with theory.

Test Severity

The major components of test severity are tire force severity, pavement surface severity, and weather/temperature severity. Biard and Svetlik (Ref 22) reported that the wear index of styrene-butadiene rubber ranged from 98 to 216 (natural rubber = 100) as severity increased. Geesink and Prat (Ref 23) found that the wear rating order of different tread compounds changed with test severity. Veith (Ref 12, 13) systematically studied tread wear as a function of the three severity factors listed above.

Tire Force Severity. Force severity was controlled by different vehicle cornering and braking maneuvers. Miller *et al.* (Ref 24) have shown that the cornering wear rate is about seven times higher than the braking wear rate. These results confirm the previous assertion that cornering force is the most important variable affecting tire wear.

Veith (Ref 12, 13) varied tread composition by using different types of carbon black and determined the tread index as a function of cornering force. Severity index gradient was defined as the rate of change of wear index with force severity. This experiment led to the conclusion that carbon black of smaller particle size provides good wear resistance at high severity conditions. Studebaker (Ref 25) and Dannenburg and Amon (Ref 26) also support this finding. As radial tires became popular, the wear rating of radial construction versus bias construction was measured. It was found that as force severity increases, the radial construction has a higher wear index than the bias construction. In other words, radial tires have higher wear resistance than bias tires.

Pavement surface severity is a combination of micro- and macrotexture and is qualitatively ranked as harsh or blunt. Lowne (Ref 27) has shown that pavement texture is important in determining wear rate. The wear ranking of tread compounds is not independent of the surface texture of the pavement. For example, a compound that is superior to another compound on a blunt surface can become inferior on a harsh surface. Veith (Ref 13) observed similar ranking reversals in testing of two different tread compounds on blunt and harsh surfaces. This sort of ranking reversal has also been observed in laboratory abrasion experiments by Schallamach (Ref 28), Kragelski and Nepomnyashchi (Ref 29), and Pulford (Ref 30). Schallamach explains this behavior by invoking the tensile tear mechanism and/or physical property degradation of rubber, depending on the nature of the abrading surface.

Weather Severity. Tire surface temperature, ambient temperature, rain, and snow affect wear. Wear rate increases as temperature increases. Veith has studied this phenomenon extensively (Ref 13, 16). The literature presents conflicting results with regard to wear rate on wet versus dry pavement.

Effects of Various Factors: Experimental Design Study

Veith (Ref 14) conducted and extensive study of the effects of multiple levels of five design factors on tread wear (Table 2). The design details, the tread wear testing, and the regression analysis are presented in Ref 14. The factors in order of importance for superior wear resistance are tire type, tread loss modulus, and tread pattern. Radial tires had better wear resistance. Similarly, as the loss modulus increased, wear resistance increased. Tires with higher groove volume and aspect ratio showed poor wear resistance.

Table 2 Effects of tire-related factors on tire wear rate

Parameter	Level	Transformation of level		Change in wear rate[a]
		From	**To**	
Tire type (ply)	Bias	Bias	Radial	-100
	Bias belted			
	Radial			
Aspects ratio	60	60	78	+39
	70			
	78			
Groove void fraction	0.16	0.16	0.43	+46
	0.43			
Tread pattern	Circumferential	Circumferential	Block	-7
	Block			
Tread loss modulus	0.84 MPa (122 psi)	0.84 MPa (122 psi)	1.37 MPa (200 psi)	-15
	1.37 MPa (200 psi)			

(a) Minus sign represents an improvement in wear rate; plus sign represents a decline in wear rate:

Veith and Chirico (Ref 13) also studied carbon black in a separate experiment. At high severities, carbon blacks with high structure and surface area exhibited better wear resistance than those with normal structure. The rate of wear passed through a minimum as a function of carbon black level at each severity. Wear rate increased with increased oil content. Some crossover results were obtained as the severity level changed from high to low.

Tire Conceptual Model

A tire is a composite deformable structure. Under load, each tire component deforms. The magnitude of these deformations varies from component to component, depending on component stiffness. The different deformations generated in each component cause interfacial stresses, leading to a complex stress distribution in the footprint area. Oblizajek and Veith (Ref 15) discussed a detailed experimental setup to measure the footprint stress distribution and mechanical properties of tire components.

The mechanical properties discussed in Ref 15 can be used to characterize wear using the Gough model (Ref 11, 31). The Gough model considers the tire tread band as a beam and the carcass as an elastic foundation. Tread/belt transverse flexural rigidity. *EI,* and carcass foundation stiffness, *K,* are introduced in this model. The flexural rigidity is a product of Young's modulus, *E,* and the moment of inertia, *I.* The effects of different combinations of *EI* and *K* values are discussed in Ref 11. It is reported that tire wear improves as the pairs of *EI* and *K* values change in the following order: high *K,* low *EI;* high *K,* high *EI;* low *K,* high *EI.* Using composite theory, Walter *et al.* (Ref 32) developed a method to calculate the tire component stiffness parameters in different directions.

Further refinement of the Gough model introduced tread band transverse shear modulus, *G,* as an additional important parameter to explain wear resistance. The modified tread band transverse stiffness incorporates both *EI* and *G* (Ref 31). Oblizajek and Veith (Ref 15) have shown very good correlation between these mechanical properties and tread wear over a wide range of tire construction.

Coefficient of Friction of Rubber. Providing a specific value for the friction coefficient of rubber is impossible because there is no one value for the coefficient of friction of rubber against a specific surface (asphalt, concrete, glass,

and so on). Friction is dependent on a complex system of parameters (that is, normal force, sliding and rotating speed, surface roughness, surface lubrication, and temperature). Grosch (Ref 33) has conducted extensive experimental studies of rubber friction over a wide range of parameters and conditions. It is postulated that the total frictional force, F_t, is given by the equation:

$$F_t = F_a + F_h + F_v \quad \textbf{(Eq 8)}$$

where F_a, F_h, and F_v are the adhesive, the hysteretic, and the viscous components, respectively. In addition, the magnitude of the friction force depends on the contact area. Muhr and Roberts (Ref 34) have published a review of rubber friction and its relation to wear. Despite all of the experimental work done or rubber friction to extrapolate data from the laboratory results, determining the actual tire tread wear rate in the field is still not easily possible.

Future Outlook

Tire tread wear is a result of dissipation of frictional work generated in the footprint area. Cornering force is the primary factor responsible for tire wear. Reasonable correlation between tire component mechanical properties and tread wear has been established by statistical methods. However, this approach has its limitations. Although much research has been directed toward gaining an understanding of tire wear, most of it so far has been empirical. More research is needed to establish a precise relationship among composite properties, material properties, and tire wear.

References

1. A. Schallamach, *Rubber Chem. Technol.,* Vol 41, 1968, p 209
2. K.A. Grosch, *Rubber Chem. Technol.,* Vol 65, 1992, p 78
3. K.A. Grosch, *Fractography of Rubber,* A.K. Bhowmick and S.K. De, Ed., Elsevier, 1991, p 139
4. C.T.R. Pulford, *Rubber Chem. Technol.,* Vol 58, 1985, p 653
5. A.H. Muhr and A.D. Roberts, Paper presented at the Rubber Division ACS Meeting (Toronto), 21-24 May 1991
6. A. Schallamach, *Trans. Inst. Rubber Ind.,* Vol 28, 1952, p 256; *J. Polym. Sci.,* Vol 9, 1952, p 385
7. K.A. Grosch and A. Schallamach, *Trans. Inst. Rubber Ind.,* Vol 41, 1965, p 81
8. A. Schallamach and K.A. Grosch, *Wear,* Vol 4, 1961, p 356
9. A.N. Gent, and C.T.R. Pulford, *J. Appl. Polym. Sci.,* Vol 28, 1983, p 943
10. V.E. Gough, *Trans. Inst. Rubber Ind.,* Vol 32, 1956, p 27
11. V.E. Gough, *Wear,* Vol 2, 1958, p 107; Paper 667A, presented at SAE Meeting, Detroit, 19-21 Mar 1963
12. A.G. Veith, *Rubber Chem. Technol.,* Vol 46, 1973, p 801, 821
13. A.G. Veith and V.E. Chirico, *Rubber Chem. Technol.,* Vol 52, 1979, p 748
14. A.G. Veith, *Tire Sci. Technol.,* Vol 14, 1986, p 201, 219, 235
15. K.L. Oblizajek and A.G. Veith, *Tire Sci. Technol.,* Vol 14, 1986, p 264
16. A.G. Veith, Paper presented at the Rubber Division ACS Meeting (Toronto), 21-24 May 1991
17. "Vehicle Dynamic's Terminology," J670e, *1992 SAE Handbook,* Vol 4, *On-Highway Vehicles and Off-Highway Machinery,* Society of Automotive Engineers, p 34.246
18. A. Schallamach and D.M. Turner, *Wear,* Vol 3, 1960, p 1
19. D.I. Livingston, Paper presented at the Rubber Division ACS Meeting (Mexico City), May 1989
20. D.I. Livingston, Plenary Lecture, Tire Society Meeting, University of Akron, 15 Mar 1991
21. A.G. Veith, *Tire Sci. Technol.,* Vol 14, 1986, p 139
22. J. Biard and F. Svetlik, *Rubber Chem. Technol.,* Vol 26, 1953, p 731
23. H.A.O.W. Geesink and L. Prat, *Rubber Chem. Technol.,* Vol 31, 1958, p 166
24. R.F. Miller, R. Marlowe, and J.L. Ginn, *Rubber Plast. Age,* Vol 42, 1961, p 968
25. M.L. Studebaker, *Rubber Chem. Technol.,* Vol 41, 1968, p 373

26. E.M. Dannenburg and F.H. Amon, *Rubber World,* Vol 131, 1955, p 627
27. R.W. Lowne, *Rubber Chem. Technol.,* Vol 44, 1971, p 1159
28. A. Schallamach, *Rubber Chem. Technol.,* Vol 31, 1968, p 209
29. I.V. Kragelski and E.F. Nepomnyashchi, *Wear,* Vol 8, 1965, p 303
30. C.T.R. Pulford, *J. Appl. Polym. Sci.,* Vol 28, 1983, p 709
31. V.E. Gough, *Kautsch. Gummi,* Vol 20, 1967, p 469
32. J.D. Walter, G.N. Augeropoulos, M.L. Janssen, and G.R. Potts, *Tire Sci. Technol.,* Vol 1, 1973, p 210
33. A. Grosch, The Speed and Temperature Dependence of Rubber Friction and Its Bearing on the Skid Resistance of Tires, *The Physics of Tire Traction,* D.F. Hays and A.L. Browne, Ed., Plenum Press, 1974, p 143
34. A.H. Muhr and A.D. Roberts, Friction and Wear, *Natural Rubber Science and Technology,* A.D. Roberts, Ed., Oxford University Press, Oxford, 1988, p 773

Friction and Wear of Aircraft Brakes

E.M. Tatarzycki and R.T. Webb, Aircraft Braking Systems

Introduction

AIRCRAFT BRAKES are designed to stop an aircraft through the conversion of kinetic energy into heat by the mechanism of friction. Aircraft brakes are composed of multiple disk pairs, which are commonly referred to as the brake heat sink. Unlike an automotive brake, which consists of a single piston, a single rotor, and two brake pads, the majority of aircraft brakes use full-circle rotors and stators (Fig. 1). Brake housings normally contain several pistons for applications of the normal force needed to develop the brake torque. Large commercial transport aircraft brakes must be capable of absorbing up to 135 MJ (100×10^6 ft · lbf) of energy. The high levels of torque developed to stop an aircraft require the conversion of large amounts of kinetic energy into thermal energy over a short period of time. This energy conversion process produces very high energy fluxes at the multiple friction interfaces, resulting in high temperatures and stresses in the brake heat sink.

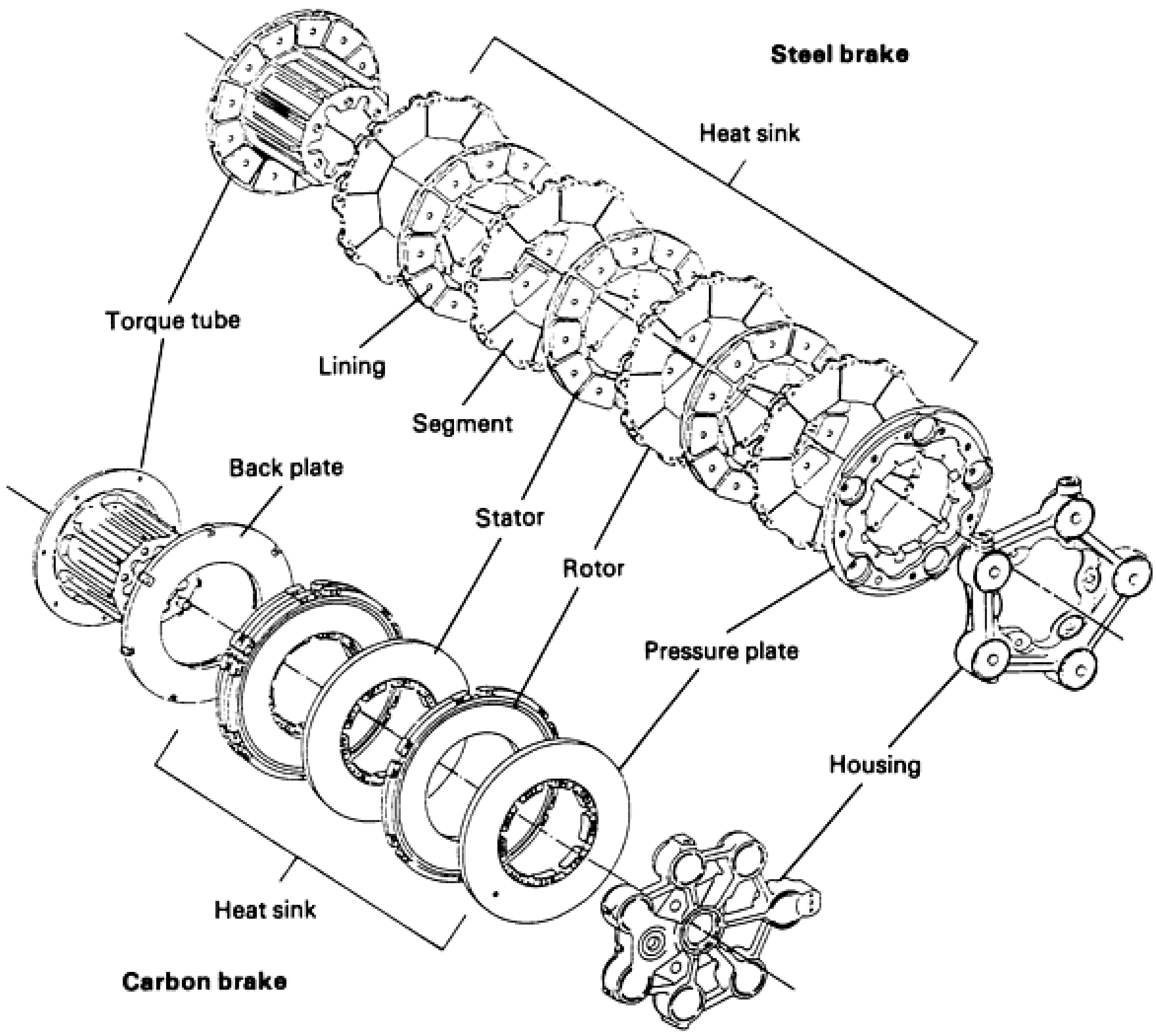


Fig. 1 Illustrations of steel and carbon aircraft brakes

The friction and wear characteristics of the friction materials used in aircraft brakes are influenced by internal factors (such as friction-material composition and heat sink mass) and external factors (such as the amount of kinetic energy absorbed by the brake, the surface velocity of the friction interfaces, and aircraft deceleration requirements). Simply put, these factors control the temperature at the interface and the normal and tangential forces of the friction material.

Aircraft Friction Materials

Three basic friction materials are presently being used in the aircraft brake. These include:

- Sintered metal friction materials
- Carbon-carbon composites
- Organic materials

The sintered metallic friction material is probably the most widely used in aircraft brakes. The steel brake heat sink consists of a sintered metallic friction material bonded to a steel supporting backing plate. The steel brake has a higher wear rate, is heavier, and has a lower cost per brake landing compared to the carbon brake. The carbon/carbon composite friction material is the latest technology. The carbon brake is lighter in weight, has excellent high-temperature performance, has a low wear rate, and has a higher cost per brake landing. Organic brake linings are used on older and

smaller aircraft and in helicopter rotor brakes. Most of the work being done in this area is aimed at replacing the existing asbestos-based linings with nonasbestos materials and will not be discussed in this article.

Steel Brakes

The metallic friction materials are primarily fabricated into two final forms. The current designs dictate that the metallic friction material be bonded into metal cups. These linings, as they are commonly called, are then riveted to a carrier disk. The other form is called a full-sintered disk on which the friction material is bonded directly to the carrier disk.

The steel brake is composed of multiple pairs of rotors and stators (Fig. 1). In this type of brake, the stators carry the metallic friction material and the rotors are composed of a high-strength high-temperature steel, commonly called the opposing surface. Some designs have the rotors carrying the metallic friction material. The opposing steel typically will last two to three times longer than the metallic friction material lining. The selection of the metallic friction material influences the brake design and must be carefully considered in order to obtain optimum dynamic performance, friction coefficient, stability, and wear rate of the friction pair. The system dynamics, which includes brake, wheel, tire, strut, and hydraulics, is also greatly influenced by the selection of the friction material.

Chemistry. The metallic friction materials are blends of various metallic and nonmetallic powders. The mixtures are primarily blends of copper and iron with one or the other being predominate in the composition. The base matrix of either iron or copper is modified by the addition of graphite, in natural or synthetic form; silicon as an abrasive; and modifiers such as high-temperature lubricants (for example, molybdenum disulfide). The size and shape of the powders can influence the final product. Proprietary materials may be added to control brake vibration, matrix strength, and wear.

Processing. The metallic mixture is then sintered in a furnace to form a metal-matrix composite. The sintering process takes place at high temperatures and sometimes under pressure. The required final density of the metallic matrix is attained by hot or cold working of the lining. The choices made in the processing of this matrix can greatly affect the dynamic performance and wear life of the friction pair. The bond of the friction material to the steel cup or core is critical. If the sintered matrix does not bond, then the matrix can spall and leave large areas void of friction material. The bonding of the matrix to cup or core is controlled by the base metallic elements (iron and copper) and any devices that might be added to the design to enhance matrix-to-carrier bond strength.

The copper-base metal matrix can attain a good bond to the carrier if sintered at an optimum temperature, time, and pressure. The iron-base metal matrix sometimes requires a screen welded to the cup bottom to supplement bonding. The iron-base metal matrix relies on a mechanical bond occurring in the open area of the screen.

Friction Material Selection. Choice of friction material determines the friction coefficient and the brake design. The iron-base matrix tends to have a lower coefficient. This requires the brake to have large-diameter pistons to achieve specific design goals (such as stopping distance, which is a critical parameter in aircraft certification). The copper-base matrix tends to have a higher coefficient. This would require smaller pistons and result in a lighter housing.

Thermal Properties. The thermal and strength properties are influenced by the chemistry, particle shape, and particle size of the friction material mixture. For a stator-rotor friction pair, the most significant parameter is the temperature at the friction interface. Typically, the lower the interface temperature, the better the brake wear life. This temperature is influenced by the thermal conductivity of the lining and opposing steel surface, the thickness of each disk, and other boundary conditions. It is desirable to design a friction pair such that the heat is conducted away from the interface quickly.

Balance. The correct distribution of mass among the various brake members is crucial for optimum dynamic performance and balanced disk wear. The specific distribution of mass in the brake is very dependent on the nature of the friction mix. Each interface of a multiple-pair brake does not perform equally. Typically, the greatest wear takes place on surfaces that are nearest the pistons. Therefore, the friction pairs are normally of greater axial thickness closer to the pistons. Mass balancing of the brake stack components is done so that the friction interfaces all wear out at the same time.

Brake wear is normally measured by the length of a steel pin (wear pin) extending from the pressure plate through the housing. If a brake has a 25 mm (1 in.) wear pin, this means that 25 mm (1 in.) of wearable material is left in the brake stack. Brake wear is usually expressed in units of mm/face/stop. In subscale dynamometer evaluations, weight-loss measurements are of primary importance.

The wear mechanism taking place at the friction surface is very complex and not fully understood. After the first few stops, both surfaces form a film, sometimes referred to as a layer of glaze, at the contact surface. This film is mainly an oxide, the mechanical and tribological properties of which are different than the parent materials. The film thickness varies over the surface. The combination of thickness and properties controls the resulting coefficient and wear rate of the brake.

Wear Mechanism. An aircraft brake is expected to operate over a wide range of energy inputs. These varying energies control the temperature at the friction interface. During low energy, that is, low interface temperature, the wear is characterized by abrasive wear. Abrasive wear occurs when hard particles cut grooves across the opposing surface and displace material. When brake energies are higher, the interface temperatures are also higher and an adhesive type of wear takes place. Adhesive wear occurs when surface asperities, bonded together under localized high temperatures and pressures, shear apart during sliding of the surfaces. The wear debris may become trapped between the surface and contribute to abrasive wear.

Wear Rates. Figure 2 shows wear rates based on normalized stack loading and average energy flux (AEF). For an aircraft to attain 1000 stops (a stop is equivalent to a landing), assuming a five-pair brake with a 50 mm (2 in.) wear pin, wear rates of 0.0025 mm/face/stop (0.0001 in./face/stop) would be required.

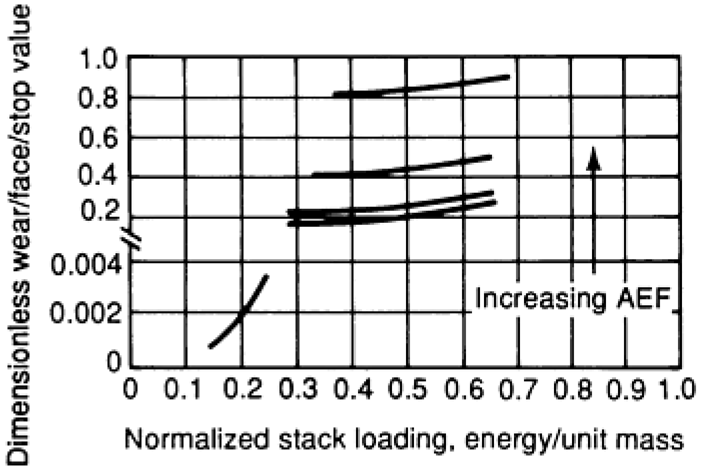


Fig. 2 Typical wear rates for a steel brake based on stack loading and average energy flux (AEF)

When aircraft taxi out to the runways, they normally start and stop along the way. These small stops are referred to as taxi stops and snubs. During a snub an aircraft does not come to a complete stop. Brake life typically is measured by the number of landings, and this would include any taxis and snubs required to maneuver the aircraft. For steel brakes, the wear that takes place at these low energies is not as significant as the main landing wear and is normally considered as being part of the main landing stop. In contrast, the taxi wear in carbon brakes accounts for a large portion of the brake wear.

Aircraft brakes are designed to handle one rejected take-off (RTO) stop (the wear rate is normally one hundred to one thousand times greater during an RTO stop than during a service stop). After an RTO stop, the brake and wheel are normally scrapped. In recent years the worn RTO has become an important issue. Because RTO wear on a worn brake stack in normally at a higher rate than wear on a new stack, some aircraft have been mandated by the Federal Aviation Administration (FAA) to limit the brake life to ensure that there is sufficient wear material in the brake to perform a worn RTO.

Brake Design. The selection of the metallic mix can influence the brake design in many ways. A friction material with a low friction coefficient will require larger pistons to supply a large clamping force (normal force) in order to achieve the high torque required to attain stopping distances. The required heat-sink mass is normally determined by how much energy the brake must absorb during the RTO stop or by some other critical condition. Typically, new brakes with metallic linings are designed to be loaded in the 750,000 to 900,000 J/kg (250,000 to 300,000 ft · lbf/lb) range for the RTO stop. The required aircraft deceleration is also influenced by mix selection. The deceleration rate controls the

average energy flux at the interfaces, and different types of mixes have differing limitations. The limitations on AEF determine the amount of rubbed area required in the brake.

Coefficient of Friction. In general, metallic friction materials exhibit a consistent coefficient of friction over a broad range of loadings, whereas carbon friction materials exhibit a wider band of coefficient over the operating spectrum of the brake. The balance coefficient is influenced by the past service history of the friction material, the loadings, and the design of the opposing surface. Normally, the brake coefficient decays during the first several landings when starting with new friction material. This decrease in coefficient is usually attributed to formation of oxides and stabilization of the surface film.

The friction coefficient is also affected by the energy flux and the total loading on the brake (Fig. 3). If a brake design requires a high energy flux, the coefficient must then be accounted for. Normally, as the energy flux is increased, the average coefficient decreases, and thus the brake clamping forces must be of sufficient magnitude to ensure that the required deceleration rate is met.

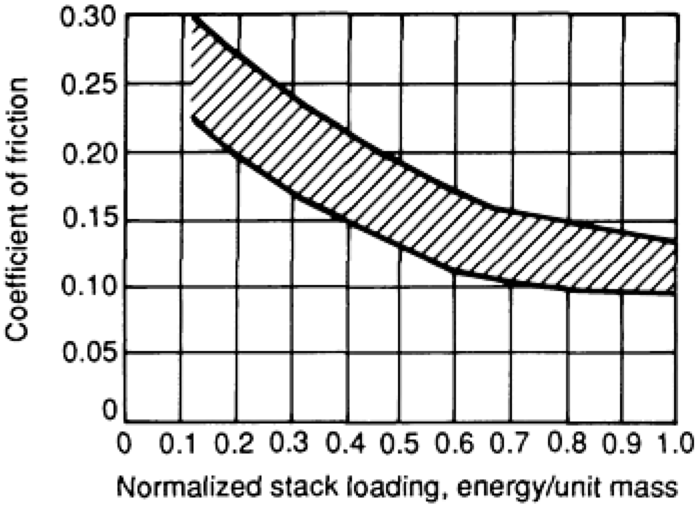


Fig. 3 Typical range of friction coefficients for a steel brake based on stack loading

The brake loading is another parameter that changes the average coefficient during a stop. The requirements of the RTO stop normally influence the brake design, and lower friction coefficients due to higher loading and surface velocity must be considered. Coefficients are also influenced by the material and design of the opposing metal surface. The majority of the steel brakes use AMS 6385C (plate), AMS 6302C (forging), or Timken 17-22AS for the opposing surface.

Spalling. The metallic friction material is exposed to high stresses and temperatures. During the extended use of an aircraft brake, the linings wear thinner and the friction material oxidizes; both factors potentially lower the matrix strength of the friction material. The high stresses encountered during stopping may spall all the friction material. The loss of large chunks of friction material reduces the rubbing contact area, thereby reducing brake torque. Once a lining spalls, the metallic cup that holds the friction material can warp. In severe cases, the rivet could fail and deline the face of one disk. In order to eliminate the potential for matrix spalling, minimum lining thicknesses have been established and friction materials with enhanced matrix strength have been developed.

Vibration. Typically, a metallic friction material that has a longer wear life also tends to cause brake vibration. It has been determined that a friction material with an increasing coefficient during a stop will tend to make the braking system unstable and cause vibration (Fig. 4). Proprietary additives and changes in the basic constituents of the friction material are made to produce a "smooth" friction material. Severe brake vibrations can cause the strut to vibrate, possibly leading to strut failure; therefore, such vibrations must be accounted for and remedied.

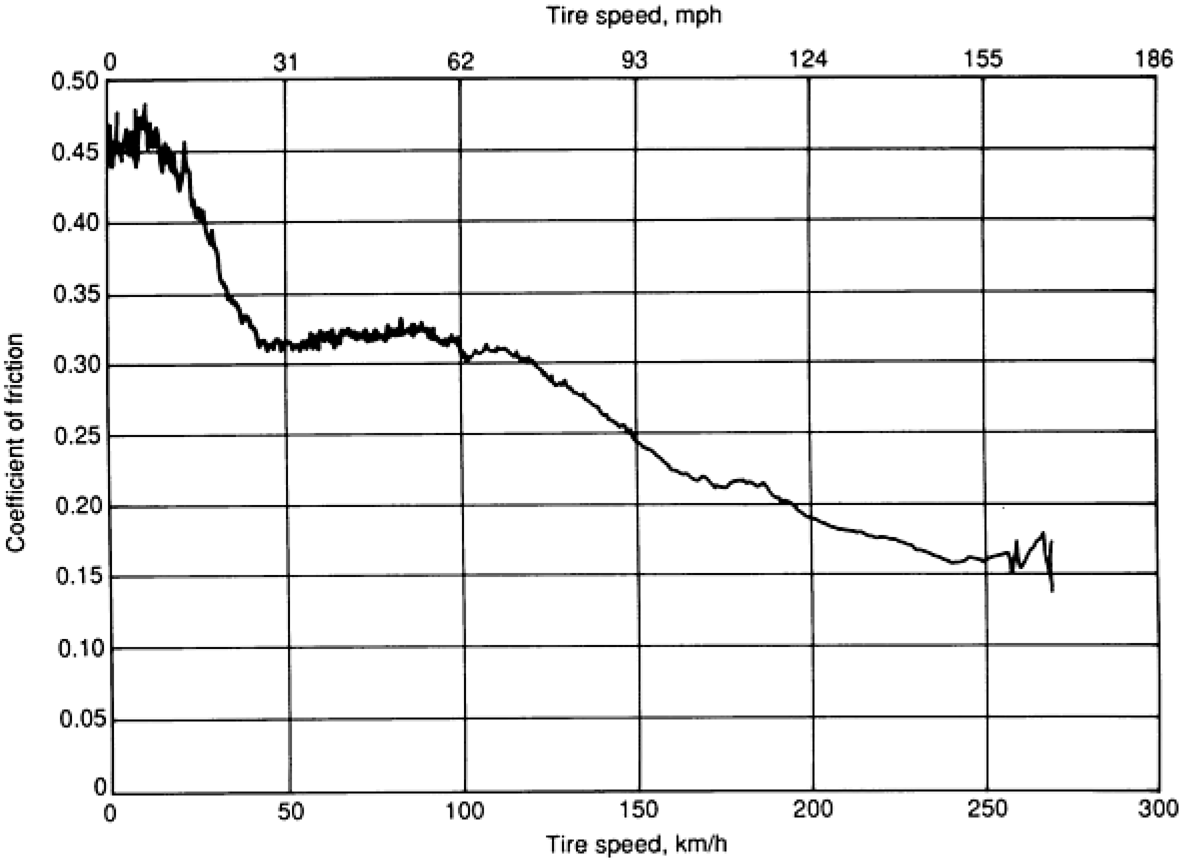


Fig. 4 Variation of friction coefficient versus tire speed during a single stop

Service Life. The aircraft industry is constantly striving to produce longer wearing brake materials. Typically, large commercial jets get around 1000 landings on a brake stack before refurbishment. The individual brake components should be balanced so that they wear evenly. Uneven wear can cause problems, such as spalling and delining.

Carbon Brakes

Raw Materials. Carbon-carbon friction materials are composites comprised of high-density carbon fibers embedded in a carbon matrix. At the present time, all the carbon fibers used in carbon aircraft brakes are made from two precursors: polyacrylonitrile (PAN) or pitch. Fiber properties are normally controlled by the manufacturing process of the fiber and are beyond the scope of this article. Such information, however, can be found in the articles "Carbon Fibers" and "Processing of Carbon-Carbon Composites" in *Composites,*Volume 21 of *ASM Handbook.*

Carbon fibers can be woven, knit, spun, and generally handled like most textile threads, but with more difficulty. In brakes, woven fabrics, short-length yarns, chopped fabrics, and woven three-dimensional preforms are used. Each manufacturer has its own preferences; thus the entire system--reinforcement and matrix--can be designed to produce specific performance.

The carbon matrix can be established in many ways. The most common is by the deposition of pyrolytic carbon directly onto the surfaces of the pores and voids within the disk. This technique is called the chemical vapor infiltration (CVI) or chemical vapor deposition (CVD) process. The infiltration or deposition is carried out in a reactor (furnace) under vacuum, high temperature, and the flow of a hydrocarbon gas. Another way to achieve a carbon matrix is to mold the fibers with a resin, such as phenolic, and convert the resin to carbon by charring at temperatures ranging from 540 to 815 °C (1000 to 1500 °F). Subsequent impregnation as with resin, pitch, or resin-pitch blends followed by a char cycle can also be carried out.

Processing. Carbon-carbon composites are porous because of the nature of the densification process. The conversion of phenolic resin to char results in porous, glassy carbon. The CVI process relies on the ability of the gas to infiltrate the maze of fibers or partially densified composite to deposit a carbon, which some researchers have categorized into three groups: isotropic, smooth laminar, or rough laminar. As the infiltration proceeds, many of the interior surfaces become inaccessible to the gas. These remain inaccessible and become permanent voids. Consequently, a multiple-step densification cycle is necessary to reduce the porosity and to achieve acceptable densities, usually in the 1.70 to 1.85 g/cm^3 range.

Processing may also include heat treatment at temperatures ranging from 1650 to 2800 °C (3000 to 5000 °F). Such a heat treatment is called graphitization and is used to control the properties of the carbon. For example, a fully graphitized carbon (graphite) will have a crystalline form, have an atomic lattice spacing of 0.3359 nm (3.359 Å), offer little resistance to indentation (making it soft), have a high coefficient of friction, and will exhibit high wear. An ungraphitized carbon will be amorphous in nature, have an atomic lattice spacing of 0.344 nm (3.44 Å), show high resistance to indentation (making it hard), produce a low coefficient of friction, and exhibit low wear. By selecting the appropriate raw material constituents, quantity of each constituent, specific process, and heat treatment temperature, it is possible to engineer a carbon for a range of friction coefficients and wear rates.

Physical Mechanical, and Thermal Properties. Carbon-carbon composites used in brake applications are unique because a single material serves as the friction surface, the heat sink, and the structural member of the brake. Consequently, this family of materials must have a wide range of properties. Densities must be high enough to enable the brake to absorb the high kinetic energies expended by the largest transports, yet low enough to be considered as a light-weight aircraft material. Its surfaces must produce a suitable rubbing aggressiveness to stop a large aircraft, yet maintain uniformity in friction over a wide range of temperatures. As a structural member, the carbon must exhibit high strength at both room and elevated temperatures.

The following properties are considered to be important for carbon brake performance:

- Density
- Porosity
- Thermal conductivity
- Specific heat
- Tensile strength
- Flexural strength
- Compressive strength
- Shear strength
- Impact strength

Coefficient of Friction. Many factors affect the coefficient of friction in a carbon brake. These factors can be grouped into two categories: overall brake design and actual carbon in the heat sink. In the brake design category, increasing the number of rotors will increase the coefficient as will a smaller size piston. The influence of carbon on the coefficient is complex. Studies conducted by the brake manufacturers have remained proprietary, but some generalities can be discussed. The PAN-base fibers will produce a different coefficient than pitch-base fibers. How different these coefficients will be is determined by processing factors. Fiber direction also determines the magnitude of coefficient. Some studies have shown that fibers in the circumferential direction produce a higher coefficient. The hardness or softness of the matrix also changes the coefficient; a harder matrix results in a lower coefficient.

The requirements for the RTO normally determine the adequacy of the brake design. A sufficiently high coefficient is needed to stop the aircraft. Proper selection of raw materials and processing conditions will determine the coefficient under RTO conditions as well as under the other energy conditions. Normally, the highest coefficient occurs at taxi conditions; the lowest occur at RTO energies. Figure 5 shows a typical variation in coefficient from taxi to RTO energies.

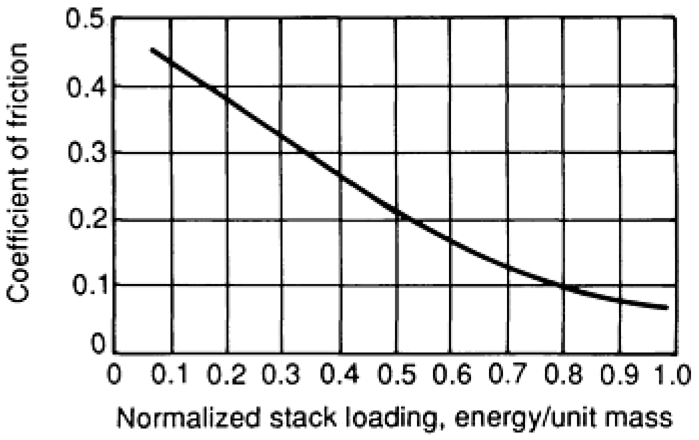


Fig. 5 Variation in coefficient of friction of carbon brake from taxi (left side) to RTO (right side) energies

Brake Wear. Carbon aircraft brake wear characteristics depend upon the raw materials and the specific processing parameters used during fabrication as well as the operating conditions of temperature, rubbing velocity, contact pressure, and prior braking history. Published information on the wear of carbon brakes is sparse, primarily because of the highly competitive nature of the aircraft brake business. Consequently, brake wear will be discussed in terms of possible underlying mechanisms rather than published scientific data.

Aircraft brakes operate under a wide range of energy conditions. Typically, an aircraft will taxi out to the runway, making several braking snubs (slowdowns) and complete stops along the way. After a successful take-off and flight, the landing occurs. Landings are classified into three typical categories:

- Service energy
- Normal energy
- Overload energy

Service energy is the lowest, and overload energy is the highest. Magnitude of energy is determined by the load the aircraft is carrying. The highest energy an aircraft will experience is during RTO.

Wear tests in the laboratory are conducted so as to simulate the conditions experienced by the actual aircraft. Taxi snubs, taxi stops, and landings are conducted in combinations called taxi/landing cycles, and wear is measured after several hundred of these cycles. Thickness loss as well as weight loss are measured. This technique enables engineers to compare different materials in the same brake run under the same conditions. In order to understand wear in engineering terms, it is necessary to run special tests at single energies. For example, 100 stops may be run at taxi energies, 100 more at service energies, 50 more at normal energies, and 20 at overload conditions. After each condition, the brake is disassembled and wear measurements taken. This approach enables the determination of which conditions result in the highest and the lowest wear. Figure 6 shows such data incorporated into a curve. Unlike a metallic brake, a carbon brake experiences the highest wear during the taxi condition, the lowest during the service condition.

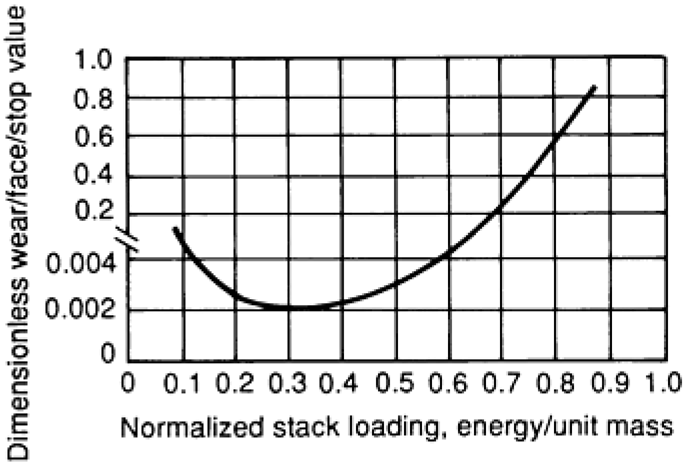


Fig. 6 Variation in wear of carbon brake from taxi to RTO energies

Each of the three energy ranges has its own wear mechanism. The taxi region exhibits the highest wear, and mechanical abrasion predominates. In the mid-energy (service energy) region, wear is lowest because a thin film protects the substrate. The third region occurs when temperatures are sufficiently high to produce oxidation of any debris or substrate (fiber or matrix). Oxidation usually occurs during overload and RTO conditions.

PAN-base fibers wear differently than pitch-base fibers; these differences are associated with fiber processing as well as composite processing. A harder matrix will wear less than a softer one. Studies have shown that circumferentially oriented fibers have high wear and a high coefficient of friction.

Wear Mechanism and Wear Debris Analysis. The worn surfaces of the carbon materials have bands of different reflectance when observed via macro examination. Debris material occurs in both bright and dull bands. The observed banding on the wear surface of the composite is related to variations in reflectance; these variations are thought to be related to differences in the character of the wear debris. The bright bands consist of a thin film of debris with a polished appearance that produces a high reflectance. The dull bands have a high number of fissures or shallow grooves and scratches in the longitudinal fiber bundles, both of which are indicative of fiber removal caused by abrasion. The debris material in the dull bands does not produce a high reflectance and appears less dense (more porous) than the bright bands. The debris in the dull bands is particulate and is comprised of fibers and matrix. This debris does not form a film, but fills in the original porosity, mainly in the matrix, and is found primarily under low-energy conditions.

The debris in the bright band consists of a thin film (up to a few micrometers thick) that is smeared over fibers and matrix and also fills in the pores. The bright bands appear to be denser (less porous) than the dull bands. The wear debris film is grooved, whereas the fibers and matrix are not. This indicates that the wear debris film is protecting the fibers and matrix from abrasion.

The thin film of wear debris is amorphous in character and is comprised of both fibers and matrix. High-energy braking conditions produce a high coverage of wear debris (film), which may result in a low coefficient of friction and low wear rate. More detailed information on wear debris analysis can be found in the article "Lubricant Analysis" in this Volume.

Moisture Problems. Carbons and graphites have an affinity for moisture. These materials adsorb moisture; that is, water molecules are attracted to their exterior and interior surfaces. Consequently, the presence of moisture significantly reduces the coefficient of friction. When a brake is sitting for a period of several hours, the rubbing surfaces will adsorb moisture from the air. Thus, when a brake stop is initiated in this condition, the stopping power is significantly reduced because of the low coefficient. This is typically referred to as "morning sickness." As rubbing continues through the stop, the moisture evaporates and the coefficient returns to its normal dry value.

Oxidation. Carbons and graphites are also subject to oxidation at elevated temperatures. Typically, the threshold of oxidation is considered to be 430 °C (805 °F). Technically there is oxidation at this temperature, but it is so low that it is considered to be negligible. Other than taxi conditions, the operating temperatures of a carbon brake will range between

500 and 1000 °C (930 and 1830 °F) during landings; under RTO energies the temperatures exceed 1300 °C (2370 °F) and oxidation becomes a significant factor. In order to keep the disks from deteriorating, an oxidation inhibitor is applied to the outer and inner diameters of each carbon brake disk. Inhibitor is not applied to the wear surfaces because it would alter the friction characteristics of the carbon.

Friction Coefficient Variability. As stated previously, friction coefficients have a larger range over the operational spectrum of the aircraft when compared to the friction coefficients for steel brakes. This coefficient range has implications for brake control (antiskid) system design. In the carbon friction material, friction coefficients can vary by factors of 3 or more over the operation range of the brake. Therefore, the brake torque can vary by a factor of 3 or more. Carbon processing and brake frame design must take this friction coefficient variation into account.

Vibration. An aircraft wheel and brake is a system with multiple degrees of freedom that is subjected to high dynamic loads. This high dynamic loading is transient in nature and can be an exciter of vibration.

In a carbon brake, the most critical vibration mode is known as whirl. This mode consists of an accordian-type action of the brake disks combined with an orbiting motion of the brake structure about the axle (Fig. 7). The torque output and friction coefficient of the brake, as functions of velocity and brake pressure, are significant parameters in determining whether or not whirl motion will occur and, if it does occur, the severity of the vibration. Severe whirl can result in damage to the carbon brake disks and other wheel/brake hardware.

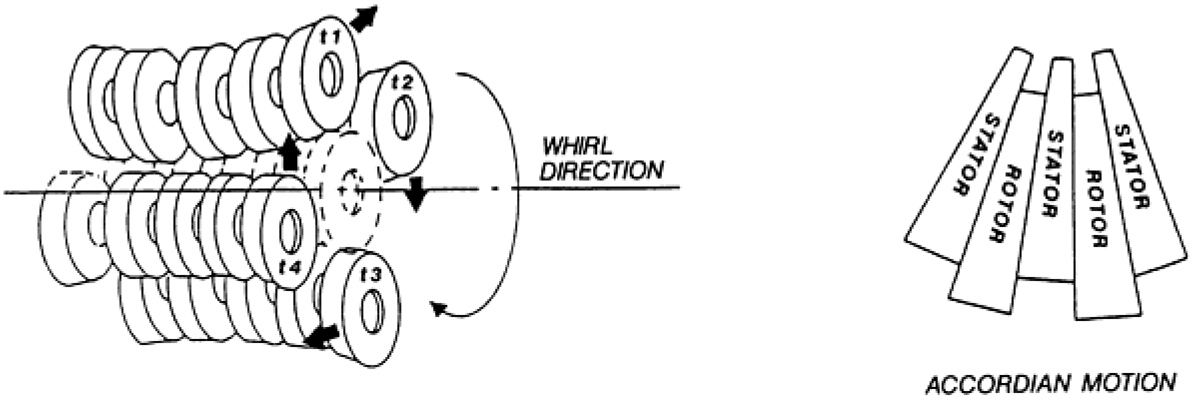


Fig. 7 Carbon brake vibration

There are two basic approaches that can be taken to control whirl vibration. First, carbon friction coefficient characteristics are engineered in such a way that the brake coefficient, at a given velocity and brake pressure, remains low enough to minimize the potential for the whirl vibration. Second, the aircraft wheel/brake is designed to incorporate the required stiffnesses and damping characteristics. Adequate stiffnesses and damping, combined with favorable friction coefficient characteristics, will ensure the stability of the aircraft wheel/brake structure at given dynamic conditions.

Testing

Friction material testing is conducted on either direct-connected dynamometers or on landing wheel dynamometers. Typically, new friction materials are screened in subscale brakes on direct-connected dynamometers. Once a potential material has been selected, then full-scale brakes are tested on landing wheel dynamometers which use an aircraft tire and wheel. Full-scale brake testing is very expensive and therefore limited. During aircraft wheel and brake qualification testing, the full-scale brake is run through numerous tests before any aircraft testing is done.

Test Requirements. Generally an aircraft brake must pass testing standards set up by the military or FAA. The military requirements are outlined in MIL-W-5013, and the FAA requirements are continued in TSO-C26. The Society of Automotive Engineers (SAE) also publishes an Aerospace Recommended Practice (ARP) for aircraft wheels and brakes (ARP 597). The airframe manufacturers also specify extensive supplemental qualification requirements that must be met before the brake can be qualified for service.

Selected References

- J.F. Archard, The Temperature of Rubbing Surfaces, *Wear,* Vol 2, 1958/59
- F.P. Bowden and J.E. Young, *Proc. R. Soc. (London) A,* Vol 208, 1951, p 444
- H.W. Chang, Correlation of Wear with Oxidation of Carbon-Carbon Composites, *International Conference on Wear of Materials,* American Society of Mechanical Engineers, 30 Mar to 1 Apr 1981
- T.S. Eyre and F. Wilson, Wear of Grey Cast Iron Under Unlubricated Sliding Conditions, *ASME/ASLE International Lubrication Conference* (New York), 9-12 Oct 1972
- D.B. Fischbach and D.R. Uptegrove, Oxidation Behavior of Some Carbon/Carbon Composites, *13th Biennial Conference on Carbon* (Irvine, CA), 1977
- K. Gopinath, G.V.N. Rayudu, and R.G. Narayanamurthi, Friction and Wear of Sintered Iron, *Wear,* Vol 42, 1977, p 245-250
- B. Granoff, H.O. Pierson, and D.M. Schuster, Carbon-Felt, Carbon-Matrix Composites: Dependence of Thermal and Mechanical Properties on Fiber Volume Percent, *J. Compos. Mater.,* Vol 7, Jan 1973
- T.-L. Ho, "Development and Evaluation of High-Energy Brake Materials," Ph.D. thesis, Rensselaer Polytechnic Institute, 1974
- J.M. Hutcheon and M.S.T. Price, The Dependence of the Properties of Graphite on Porosity, *Proceedings of the Fourth Conference on Carbon,* Pergamon Press, 1960
- W.V. Kotlensky and P.L. Walker, Jr., Crystallographic and Physical Changes of Some Carbons Upon Oxidation and Heat Treatment, *Proceedings of the Fourth Conference on Carbon,* Pergamon Press, 1960
- I.V. Kragelskii, *Friction and Wear,* Butterworths, Washington, 1965, p 117
- J.K. Lancaster, Instabilities in the Frictional Behavior of Carbons and Graphites, *Wear,* Vol 34, 1975
- R.L. Lewis and R.E. Raymond, "Stopping Distance Analysis," Society of Automotive Engineers, Inc., Paper No. 730193, 1973
- F.F. Ling and E. Saibel, On Kinetic Friction Between Unlubricated Metallic Surfaces, *Wear,* Vol 1, 1957/58
- J. Molgaard and V.K. Srivastava, The Activation Energy of Oxidation in Wear, *Wear,* Vol 41, 1977
- N. Murdie, C.P. Ju, J. Don, and F.A. Fortunato, Microstructure of Worn Pitch/Resin/CVI C-C Composites, *Carbon,* Vol 29, 1991, p 335-342
- D. Pavelescu and M. Musat, Some Relations for Determining the Wear of Composite Brake Materials, *Wear,* Vol 27, 1974
- T.F.J. Quinn, A.R. Baig, C.A. Hogarth, and H. Muller, Transitions in the Friction Coefficients, the Wear Rates, and the Compositions of the Wear Debris Produced in the Unlubricated Sliding of Chromium Steels, *ASME/ASLE International Lubrication Conference* (New York), 9-12 Oct 1972
- E, Rabinowicz, *Friction and Wear of Materials,* John Wiley & Sons, 1965
- D.M. Rowson, The Interfacial Surface Temperature of a Disk Brake, *Wear,* Vol 47, 1978
- L. Rozeanu, Friction Transients (Their Role in Friction Failures), *Trans. ASLE,* Vol 16, 1975, p 257-266
- J.J. Santini and F.E. Kennedy, Jr., An Experimental Investigation of Surface Temperatures and Wear in Disk Brakes, *Lubr. Eng.,* Aug 1975
- P. Stanek, N. Murdie, E.J. Hippo, and B. Howdyshell, The Effect of Fiber Orientation on Friction and Wear of C-C Composites (Extended Abstracts), *Biannual Conference on Carbon* (Santa Barbara, CA), 1991, p 378-379
- I.L. Stimson and R. Fisher, Design and Engineering of Carbon Brakes, *Philos. Trans. R. Soc. (London) A,* Vol 294, 1980
- E.M. Tatarzycki, Friction Characteristics of Some Graphites and Carbon Composites Sliding Against Themselves, *13th Biennial Conference on Carbon,* 1977
- A.K. Vijh, The Influence of Solid State Cohesion of Metals and Non-Metals on the Magnitude of

Their Abrasive Wear Resistance, *Wear*, Vol 35, 1975

Wear of Jet Engine Components

J.D. Schell and K.P. Taylor, General Electric Aircraft Engines

Introduction

A JET ENGINE is a sophisticated piece of machinery with many moving parts; the potential for wear problems exists whenever moving parts come into contact or unintended motion occurs between stationary contacting parts. As for any system, wear in jet engines can be controlled through proper design, material selection, and lubrication. A schematic cross section of a typical jet engine is shown in Fig. 1. The major engine subsystems consist of the fan, the high-pressure compressor (HPC), the combustor, the high- and low-pressure turbines (HPT and LPT), and the exhaust nozzle. The engine design contains one nonrotating system and two concentric rotating systems. The nonrotating (stator) system is made up of structural frames and casings. The low-pressure rotating system consists of the fan disk(s) and fan blades, the LPT disks and turbine blades, and a connecting shaft. The high-pressure rotating system consists of the HPC disks/spools and compressor blades, the HPT disks and turbine blades, and a connecting shaft.

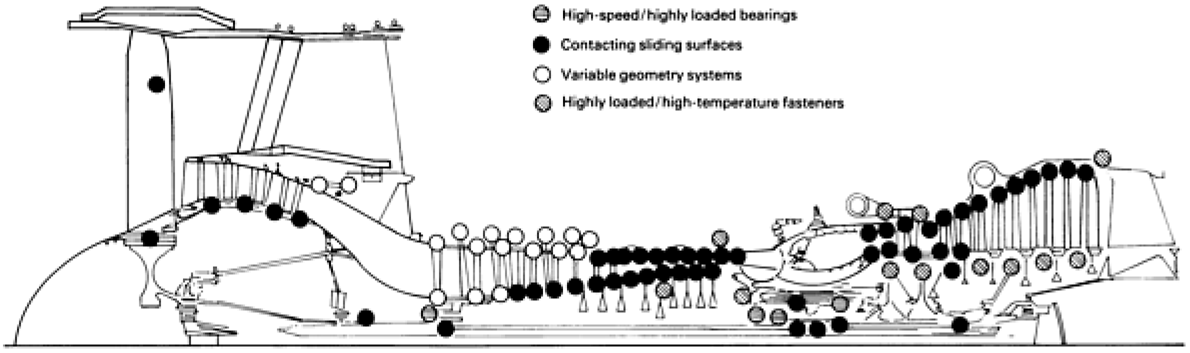


Fig. 1 Jet engine cross section showing important subsystems and potential areas of wear

Operating environments vary widely between different sections of the engine and depend on where the engine is in its mission. Temperatures may vary from subzero to above 1095 °C (2000 °F), rotational speeds may climb to more than 15,000 rev/min, and contact loads may range from a few psi to local hertzian stresses well beyond 1720 MPa (250,000 psi) in rolling-element bearings. The relative motions of components may be unidirectional sliding of rotating parts on stators, oscillatory sliding varying from a few thousandths of an inch up to several tenths of an inch, or vibratory motion resulting in impact between components. Components also may be subjected to ingested particle impacts.

The wide variety of operating conditions results in a wide variety of materials used to meet the design needs of the engine. Aluminum and titanium alloys, plastics, and resin-matrix graphite composites are frequently used in the fan and the engine nacelle. The HPC uses titanium alloys, nickel-base superalloys, such as Inconel 718, and steels, such as M152, 17-4PH, and A286. The combustor requires heat-resistant nickel or cobalt alloys, such as Hastelloy X or Haynes 188, and stainless steels for fuel tubing. The turbine sections rely on cobalt and nickel superalloys, such as Inconel X750, MAR-M-509, René 77, René 80, René 125, and advanced directionally solidified and single-crystal alloys. Often the design demands on materials for jet engine components will not permit substitution of materials for wear purposes, so a number of surface coatings and treatments are employed for wear protection.

The different operating environments and types of materials in each section of the engine result in a variety of wear types, including fretting, impact, adhesive, high-speed and oscillatory sliding, oxidational, ingested particle erosion, and abrasive wear. High-speed sliding wear occurs in rotating gas path seals. Impact wear can occur in loose part assemblies or blade midspan or tip shroud interlocks. Fretting wear is frequently seen in blade dovetails. Erosion occurs when dirt and sand particles are ingested with the air through the fan and compressor. Bearings and gears can experience rolling contact fatigue. High-temperature components can experience oxidational wear. This article will discuss some of the most significant of these wear problems in relation to specific jet engine components.

Gas Path Seals

A major area of wear in jet engines involves gas path sealing. Such seals include blade tip seals, labyrinth seals, and leaf and spline seals. Blade tip and labyrinth seal problems are concerned with clearances between rotating parts and their adjacent stators. Engine efficiency is significantly affected by the amount of gas leakage over blade tips or through labyrinth seals. In an ideal engine, the blade tips or labyrinth seal teeth would maintain minimum clearance with the adjacent stator surfaces at all points in the engine cycle. In practice, the rotor parts and stator parts experience differential growth rates because of thermal gradients in the engine and much larger mechanical growth of the rotor than the stator because of centrifugal forces. The results of these differences are depicted in Fig. 2. When the rotor and stator diameters are plotted against time after the throttle is applied for takeoff, they are seen to experience a period of interference (pinch point in Fig. 2) that causes wear. The issue can be further complicated by the stator going out of round (Fig. 3).

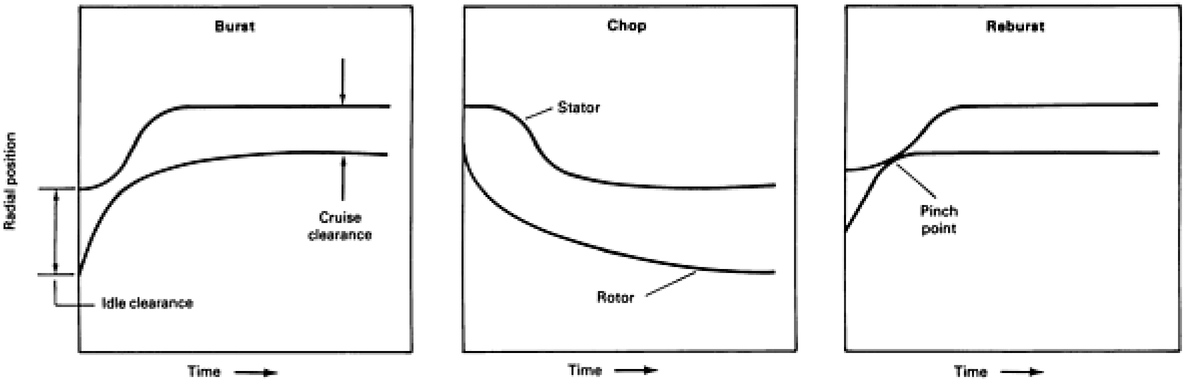


Fig. 2 Rotor and stator growth rates as a function of time and engine throttle movements

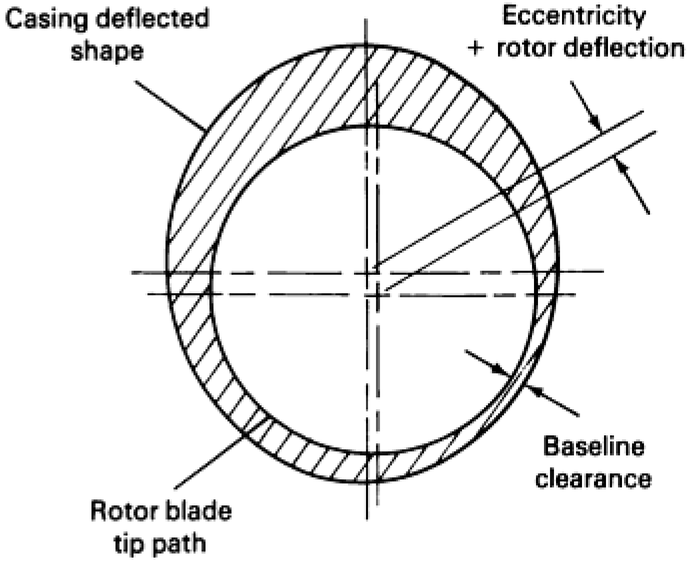


Fig. 3 Clearance change caused by rotor/stator eccentricity or maneuver deflections

Both design and material approaches are employed in combating wear of blade tips. One design approach, known as active clearance control, applies heating or cooling to the stator to achieve a better match of the thermomechanical responses of the stator and rotor. A second design approach for the out-of-round condition involves local arc grinding to remove casing material in the areas where minimum radii would occur. A materials approach that has been used in recent years is called passive clearance control. The casing is made of an alloy with a low coefficient of expansion, such as IN909, to achieve a more favorable overall thermal transient response match of the rotor and stator diameters. The most commonly used materials approach involves the application of an abradable material to the stator. The abradable material wears preferentially in a limited arc when the stator is out of round or when the rotor moves off center. This results in local clearance increases during rotor/stator interferences instead of wearing the rotor and causing a 360° clearance increase. An alternative materials approach is to apply an abrasive to the rotor, which machines the stator material, thus achieving the same result.

The materials used for the abradable stator seals or abrasive rotor coatings vary by location in the engine. Abradables can take several forms, including bonded elastomers, braze-attached sintered porous metals or honeycomb cells, or thermal spray coatings. Some of the more commonly used abradables are listed in Table 1. These materials are designed to wear in preference to the opposing blade tip or seal tooth. They rely on low densities created by included porosity or friable structures with weak bonding between constituent materials. Bill and Wisander (Ref 1) have provided a model for friable abradable seal materials. In practice, however, wear usually occurs on both surfaces, necessitating periodic overhaul.

Table 1 Commonly used abradable seal materials

Type of seal	Material
Fan and booster seals	Phenolic/carbon microballons
	Aluminum
	80/20 nickel-graphite
	Porous Teflon
	Aluminum-silicon/polyester
	Ni-Cr-Al/bentonite
High-pressure compressor seals	Nickel-graphites (75/25, 80/20, and 85/15)
	Nickel-aluminum
	Aluminum
	Aluminum bronze/nickel-graphite
	Ni-Cr-Al/nickel-graphite
	Ni-Cr-Al/bentonite
	Hastelloy X open-faced honeycomb
	FiberMetal
High-pressure turbine seals	Co-Ni-Cr-Al-Y
	Bradelloy (Hastelloy X honeycomb + braze/nickel-aluminum

The abrasive materials approach has been used with success on rotating parts, allowing them to machine their own clearances and minimizing rotor wear. The most commonly used abrasive is plasma spray aluminum oxide on seal teeth or rotor lands. Figure 4 shows a sector from an HPC rotor with two sets of seal teeth coated with plasma spray aluminum oxide. The most common mating stator seal material for such applications is open honeycomb (Fig. 5). Commonly used abrasive coatings for clearance control in jet engines include:

- Plasma spray aluminum oxide
- Entrapment-plated cubic boron nitride (Borazon)
- Entrapment-plated aluminum oxide

The abrasive coatings approach is usually combined with honeycomb or an abradable seal to improve the overall wear system for both surfaces.

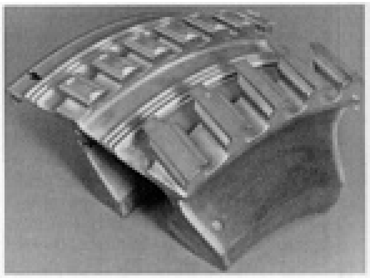


Fig. 4 High-pressure compressor disk with seal teeth coated with plasma spray aluminum oxide between stages. Arrow indicates location of coated seal teeth.

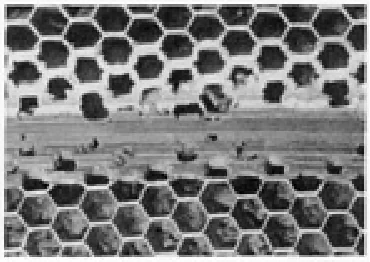


Fig. 5 Open-faced honeycomb seal showing cutting by seal teeth

Blade Midspan Stiffeners and Tip Shrouds

Some fan, HPC, and LPT rotating airfoils (blades) require the use of either a midspan stiffener or a Z-notch tip shroud (often called interlocks) to prevent mechanical flutter of the aerodynamically loaded blades. These must be designed so that the blades are sufficiently loose to allow easy assembly, but lock up into a solid stiffening ring as aerodynamic loads are imposed on the blades, causing them to untwist along the blade stacking axis. These two requirements result in a combination of impact and sliding as the interlocking contact surfaces engage and rotate into position to form the solid stiffening ring.

The impact loads imposed on the contact surfaces can be on the order of 7 to 70 MPa (1000 to 10,000 psi) and can cause severe wear damage to most materials suitable for use as blades. Therefore, it is common practice to apply a wear material to the interlock contact surfaces. These wear treatments are usually coatings on the order of 0.13 to 0.25 mm (0.005 to 0.010 in.) thick or welded hardfacing deposits up to 2.5 mm (0.100 in.) thick. Much care must be taken in the design and assembly of alignment tolerances for interlocks to prevent excessive wear, chipping, or spallation of even the most successful wear treatments on the interlock contact surfaces.

The materials used for fan and HPC blades with interlocks are usually titanium alloys, which have poor wear properties. Most fan blade and HPC interlocks use thermal sprayed WC-Co coatings or brazed-on WC-Co powder metallurgy wear pads to prevent excessive wear. The most widely used coating is Union Carbide's LW1N40, applied using a detonation gun (D-gun). Recent advances in thermal spray coatings have allowed the use of high-energy plasma spray WC-Co coatings, which hold promise for direct substitution, or high-velocity oxyfuel (HVOF) sprayed WC-Co coatings on titanium alloy interlocks. The WC-Co coatings are successful in the titanium alloy interlock applications because of the high wear resistance of the tungsten carbide, adequate fracture toughness because of the cobalt matrix, high adherence on the titanium alloy substrates, and a good match in coefficient of thermal expansion with the titanium alloy substrate materials. The typical range of temperatures for fan and HPC interlocks may vary from subzero to 95 °C (200 °F) in the fan and from about 40 to 260 °C (100 to 500 °F) in the HPC. Fortunately, WC-Co coatings appear to retain sufficient low-temperature ductility and high-temperature oxidation resistance over these temperature ranges. The formation of a wear glaze at the contact zones contributes to the good wear resistance of the WC-Co in these interlock applications.

The LPT blade materials are typically nickel-base superalloys, such as René 77 or René 125, which usually possess fairly good sliding wear resistance. However, they have inadequate wear resistance in the combined impact and sliding wear environment of LPT blade interlock contact surfaces. Typical use temperatures for LPT interlocks are 540 to 925 °C (1000 to 1700 °F), so the oxidation properties of the alloys under the existing wear conditions also play a significant role in their wear resistance. Typical wear coating compositions applied by thermal spraying or weld buildup that are used for LPT blade interlocks include:

- Tribaloy 800 (plasma sprayed, welded, HVOF)
- Cost Metal 64 (welded)
- Chromium-carbide/nickel-chromium (plasma sprayed, HVOF, D-gun)

Most of these alloys are cobalt based for good wear resistance and benefit from the formation of cobalt oxide and/or spinel wear glaze films.

Tribaloy 800 and Coast Metal 64 are the most commonly used LPT blade interlock coatings at GE Aircraft Engines. Tribaloy 800, applied by thermal spraying or tungsten inert gas (TIG) welding, provides excellent wear resistance and oxidation resistance to about 840 °C (1550 °F). Above this temperature. TIG-welded Coast Metal 64 provides better wear and oxidation resistance than Tribaloy 800. In general, wear coating performance for LPT blade interlocks correlates to the chromium content and use temperature, with better performance at elevated temperatures for coatings with higher chromium contents and better performance at lower temperatures for coatings with lower chromium contents.

Mainshaft Bearings

The materials traditionally used for gas turbine mainshaft bearings are 52100 and M50 steels. More recently, powder metallurgy (P/M) bearings and case-carburized M50NiL steel, a modified M50, have been introduced as race materials. Several factors have contributed to this recent trend. Newer gas turbine mainshafts operate at higher speeds. This has pushed the bearing DN values (bore diameter in millimeters times shaft revolutions per minute) well past 2 million, which increases race hoop stresses and the hertzian contact stresses between rolling elements and the races.

The higher hoop stresses can cause fracture of the 52100 or M50-type races because they are through-hardened materials, typically in the 50 to 60 HRC range, and thus have low fracture toughness. At high DN values, this can become a fracture reliability problem for a statistically significant number of these bearings. The higher hertzian stresses, approaching 2400 MPa (350,000 psi) for a 2.5 million DN mainshaft bearing outer race, also can cause significant reductions in rolling contact fatigue life. This is undesirable, because changing a mainshaft bearing requires costly disassembly of the engine.

Bearing races made from the new P/M alloys and forged low-carbon alloys with carburized surfaces do not have these shortcomings. These materials are designed for high DN use and require special manufacturing processes. GE Aircraft Engines has concentrated on a variation of M50 steel with reduced carbon and increased nickel to improve fracture toughness. The race is then carburized to produce a fine dispersion of carbides for high hardness. Compressive residual stresses are frozen in to the raceway surfaces to improve rolling contact fatigue, while the low hardness (<50 HRC) rare core material remains tough to deal with high hoop stresses.

This M50NiL material with its finely dispersed carbides, as well as the fine-grained (to improve fracture toughness) P/M race materials, can suffer from a low tolerance to wear. Wear can occur at the ball cage guide lands under marginal lubrication conditions even for "normal" bearing cleanliness operation. The rolling-element cage shoulders are silver plated, which provides solid lubrication and low friction to prevent wear when direct metal-to-metal cage skidding occurs on the cage guide land of the race. This works quite effectively for 52100 and M50 steels. However, the P/M and M50NiL steels sometimes experience rapid wear under similar operating parameters.

Research by Budinski (Ref 2) has shown that the size, distribution, type, and volume fraction of carbides in tool steels can significantly alter their abrasive wear resistance, with coarser carbide grains having better resistance than finer carbide grains. Thus, it has been suggested that the coarser carbide stringers in M50 or 52100 forged bearing races can adequately resist the initiation of abrasive wear, while the very fine; evenly dispersed carbides in M50NiL or P/M bearing races cannot.

The abrasive particles found in the bearings originate in the oil supply system. Sump castings, abrasively cut tubes, and grit-blast-cleaned parts in the bearing lubrication system all likely contain very fine alumina or silicon carbide contaminant particles. Most of these are removed during cleaning prior to assembly or by in-line filtering, but some of the finer particle (<50 μm) get through to the bearings even under the most stringent clean-room assembly conditions. Once inside a bearing, contaminant particles can become embedded in the silver plating on the ball cage shoulders, where they protrude, causing abrasive wear to initiate during transient cage shoulder/race guide land contacts. The abrasive particles soon become "capped" with race transfer material, and adhesive wear ensues.

The high differential sliding speeds between the orbiting cage and race cause frictional heating, local oxidation, and carburization by oil coking of the thin metallic transfer layers. Thus, a very hard abrasive transfer layer results and the wear process accelerates. These deposits also increase friction. Therefore, when cage-to-cage encounters occur, a more severe rebound force results from the skidding contact, generally increasing the cage orbiting and number of skid contacts and producing further wear.

Combustor and Nozzle Assemblies

These engine components are subjected to a variety of severe wear environments. Combustor hardware includes fuel nozzles, swirlers, and cowl damping wires, which experience relatively high temperatures (540 °C, or 1000 °F, and up) during operation. Exhaust nozzle assemblies are characterized by many parts, such as pins, bushings, links, and overlapping flaps, which aid in motion of the nozzle to control engine thrust. Some of these nozzle parts are directly in the hot gas stream and experience temperatures up to 815 °C (1500 °F); others are bathed in bypass cooling air and remain relatively cool (approximately 315 °C, or 600 °F). Combustor and nozzle assemblies experience large amounts of vibration from turbulent air flows both inside and outside the engine. This vibrational/impact wear can cause significant material removal as well as high-cycle fatigue of some components. The combined effects can cause liberation of hardware; in the case of the combustor, this will in turn cause severe damage to downstream components, such as turbine nozzles and blades.

The majority of wear problems in both of these assemblies is cause by vibration and impact. Because of the elevated temperatures in the combustor, oxidational wear occurs as scales are formed and subsequent chipped off by impact. Contact pressures between parts are nominally low, but can be aggravated by high-frequency impacts, which may locally yield the materials. Oscillatory sliding (galling) wear sometimes occurs on exhaust nozzle flaps as they are actuated during mission cycles over several hundred accumulated flight hours.

The design of these components addresses temperature and fatigue concerns. Both combustor and exhaust nozzle hardware are made from heat-resistant superalloy sheet materials, such as Hastelloy X, René 41, or Haynes 25. Because these materials vibrate in the turbulent hot gas stream, high-cycle fatigue life at elevated temperature is important. The cooler sections of the exhaust nozzle sometimes use high-temperature titanium alloys, such as Ti-6Al-2Sn-4Zr-2Mo, to reduce engine weight and maintain mechanical properties at elevated temperature. Many pins and bushings are manufactured from steel alloys, such as 17-4PH and A286. Coatings can be applied to problem areas on specific components, but they must withstand the application temperatures and not degrade the mechanical properties of the base alloy to unacceptable levels. Therefore, specification of the material and/or coating can be a complicated process.

In general, cobalt-base alloys, such as Haynes 25 and Haynes 188, tend to perform best at temperatures above 540 °C (1000 °F) in both sliding and impact wear. Some nickel-base alloys, such as René 41, also possess good wear resistance at high temperatures. Effective coatings for wear problems in these temperature regimes include Tribaloy 800 and chromium carbide/nickel chromium. Tribaloy 800 derives its good elevated-temperature wear resistance from a hard Laves phase in a cobalt-base matrix. This matrix produces a cobalt oxide, which provides lubricity to the interface. Chromium carbide/nickel chromium derives its good performance chiefly from the hard carbide phase and the formation of favorable oxide wear glazes. Cooler titanium components in the exhaust nozzle generally have poor wear resistance and almost always require coatings for mating parts in relative motion. Here, the coating of choice is generally WC-Co, which again derives its wear resistance from the hard carbide phase. Oxidation of the carbide limits use of this coating to temperature regimes below 480 °C (900 °F). Because of the aggressive nature of the carbide, both mating surfaces should be coated. The steels used in the actuation systems for the nozzle flaps are usually ion nitrided to develop a hard case layer on the surface (hardness of up to 72 HRC can be achieved). For a particularly severe wear environment, ion nitriding may not provide sufficient wear protection, and chromium carbide or tungsten carbide coatings may be required.

Dovetails

Aircraft engines utilize bladed rotors in the low-pressure fan and the high-pressure compressor to generate thrust and provide the required air compression ratios for proper fuel combustion to drive the turbine sections. These airfoils are attached to their respective disks on the rotor by dovetails (Fig. 6). As the rotor speed increases, the disk posts grow and open while the mass of these blades is forced radially outward by centrifugal forces. These forces result in relative motion between the blades and the disk under very high normal stresses, reaching localized peak loads of up to 930 MPa (135 ksi) in the fan.

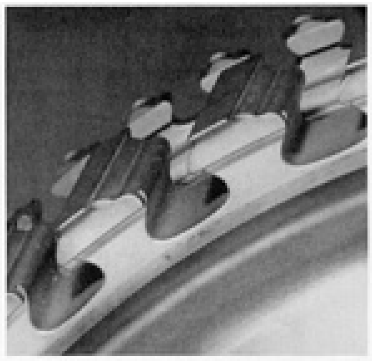


Fig. 6 Front-stage HPC rotor disk showing dovetail disk slots and pressure faces

Superimposed on some stages of the high-pressure compressor are high-frequency, low-amplitude vibrations caused by local aerodynamic effects from vane or strut positions relative to the rotating hardware. The combination of contact stress and high-frequency motion admits a large amount of mechanical work into the metal surface. Temperatures for dovetails range from subzero (fan) to close to 650 °C (1200 °F) (rear compressor). The combination of these factors creates wear problems for critical rotating hardware.

Fan and front-stage compressor blades and disks are manufactured from titanium alloys; middle- and rear-stag compressor rotor materials included steel (A286) and nickel-base alloys (Inconel 718). Rear-stage components rarely exhibit wear; however, titanium alloys have notoriously poor wear resistance when unprotected. Any cause of surface damage, such as wear, must be minimized because of the surface sensitivity of titanium in relation to high-cycle fatigue crack initiation.

Most jet engines in service use a titanium protection system comprising a thermal spray coating and a solid film lubricant on the blade dovetail. Most disk and spool dovetail slots cannot accommodate a thermal spray coating in the disk slots because they are relatively small, but the largest fan disk slots use the same coating system as the blade dovetails. The coating of choice is a relatively soft copper-nickel or copper-nickel-indium alloy applied by air plasma spraying. The thermal spray coating is subsequently dry film lubricated with molybdenum disulfide in an organic resin binder that is oven cured for strength and adhesion. The solid film lubricant provides a low coefficient of friction to prevent galling, allow part seating, and delay the onset of wear produced by metal-to-metal contact. Coating thicknesses range from 0.05 to 0.28 mm (0.002 to 0.011 in.) for the thermal spray and 0.025 to 0.075 mm (0.001 to 0.003 in.) for the cured lubricant.

Titanium galls readily in contact with other metals in relative motion, even against the soft thermal spray coating, so solid film lubricant life is important. Unfortunately, copper-nickel and copper-nickel-indium alloys and MoS_2 can oxidize above about 315 °C (600 °F), while the organic binder breaks down and loses strength at such temperatures, resulting in higher friction coefficients, increased shear forces, and wear. Improved dovetail coatings for the 315 to 540 °C (600 to 1000 °F) range need to be developed, as new high-temperature titanium alloys are being specified farther back in the compressor

Compressor Airfoil Erosion

The introduction of underwing-mounted jet engines has brought with it the problem of the ingestion of materials left on airport runways. Sand and dirt are used during winter months to improve traction on snow-covered runways. This grit is easily ingested by the engine when the snow melts and evaporates. Because the current "hub-and-spoke" airline system produces many takeoff/landing cycles in a short time, compressor blades and vanes erode rapidly from these contaminants. Erosion is also prevalent in helicopter engine compressor because of the harsh unpaved environments in

which they operate. Helicopter engines have particle separators to eliminate very large grit sizes, but finer grit sometimes enters the compressors. This particle ingestion is severe enough to cause erosion of titanium-, nickel-, and steel-base materials to various degrees.

Erosion occurs by two distinct mechanisms: high-impingement angle and low-impingement angle. High-impingement-angle erosion is characteristic on the front stages of the high-pressure compressor when large ingested particle strike the leading edges of the rotating blades. In the most severe form of this type of erosion, the leading edge can be deformed by burring or plastic deformation. Otherwise, high-angle-impingement attack usually results in blunting of the leading edge. Both changes in leading edge profile cause aerodynamic flow problems that reduce compressor efficiency and increase fuel consumption.

Low-impingement-angle erosion is characteristic on the rear stages of the high-pressure compressor. As the ingested material swirls through the compressor, it becomes fragmented and centrifuged radially toward the casing. This results in glancing impacts by angular, sharp grit near the trailing edge of the airfoil and along the tip. The loss of airfoil material causes thinning of the trailing edge and opens tip clearances near the casing, reducing compressor performance. Figure 7 illustrates the mechanism of low-impingement-angle erosion

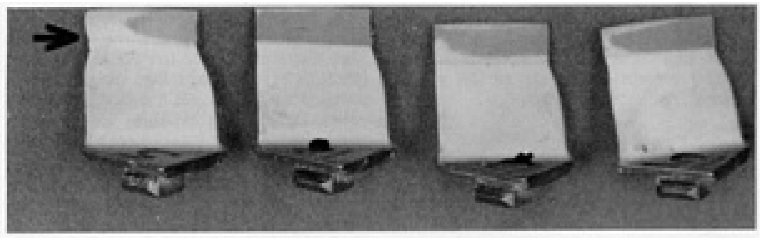


Fig. 7 Rear-stage compressor blades illustrating the low-impingement-angle erosion mechanism. Arrow indicates lost erosion coating on blade trailing edge.

Protection of individual blades has historically consisted of coating those airfoil surfaces most affected by erosion: the leading edge, across the concave surface, and back to the trailing edge in the tip region of the blade. Based on field experience, coating practice has been modified to include wrapping the coating around the leading edge and extending the coating down the airfoil surface in the direction of the dovetail. Coating materials of choice are those with higher hardness than the impinging grit, such as carbides and nitrides of titanium, tungsten, and chromium. These coatings can be applied by processes such as physical vapor deposition (PVD), D-gun, or HVOF thermal spraying to maximize their density and hardness.

The chosen coating system must not affect the mechanical properties of the base alloy, such as high-cycle fatigue resistance. Blade airfoils pass through several resonant frequencies during an engine cycle. Degradation of fatigue strength by the coating could cause cracking and liberation of the blade tip, producing more damage downstream. The coating system must also exhibit excellent adhesion to resist spallation and produce a very smooth surface finish for compressor efficiency. Any coating-induced surface defect will cause unacceptable turbulence, which can disrupt the axial flow through the compressor. Coating thicknesses are relatively thin, typically less than 0.025 mm (0.001 in.) for PVD coatings and 0.050 to 0.075 mm (0.002 to 0.003 in.) for D-gun and HVOF coatings. Coatings must remain thin so that they taper smoothly to the uncoated blade, without a sharp step that would cause compressor airflow turbulences.

Some helicopter engine compressor are constructed from one-piece blade and rotor components, called blisks. These components do not allow for the coating of individual airfoil surfaces, because the blades are an integral part of the disk (Fig. 8). Blisks also have twisted airfoils in the direction of the stacking axis, making them difficult to coat using a line-of-sight process such as thermal spray. Leading edges are extremely thin, sharp, and prone to severe burring and blunting under particle impingement. Development of non-line-of-sight erosion coatings for difficult geometries such as these is required to maintain helicopter performance.

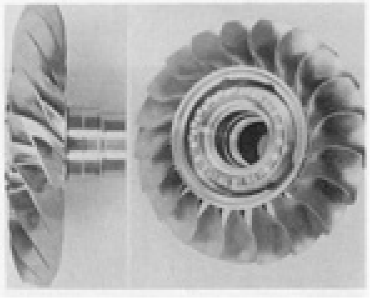


Fig. 8 Front-stage helicopter engine compressor rotor blisk

Discussion and Summary

Jet engine designs push components to their limits for mechanical strength and temperature capability. New engine designs for commercial airplanes and military fighter jets introduce new requirements for thrust and performance. The fact is that the hotter the combustion in the engine, the more powerful and efficient the engine becomes. Advances such as new superalloys, improved cooling flow designs, and thermal barrier coatings were developed to stretch the capabilities of these components and to take advantage of the thrust and efficiency gains associated with increased combustion temperatures.

A second method of generating more thrust is to increase the size of the fan and thus of the turbine required to drive it. Increased size results in larger structures with more thermal movement. Relative motion and vibration are also expected to increase with these new designs.

To accommodate size increases without substantially increasing weight, lighter materials are being developed with similar strengths to handle these more demanding conditions of temperature and loading. These advanced materials include ceramics, organic-matrix composites, metal-matrix composites, ceramic-matrix composites, and intermetallics. Each class has its own property challenges that must be overcome before it can be incorporated in engine designs. For example, composites utilize second phases to improve properties lacking in the matrix material, such as tensile strength and ductility. Intermetallic and ceramic materials can be very sensitive to any type of surface defect that may cause brittle fracture, including wear or erosion.

These advanced mechanical designs and material systems provide research opportunities in the field of tribology. The effects of increased temperature, of increased motion and vibration, and of second phases in composites on the wear process require further investigation.

References

1. R.C Bill and D.W. Wisander, *Friction and Wear of Several Compressor Gas-Path Seal Materials,* NASA TP-1128, National Aeronautic and Space Administration, 1978
2. K.G. Budinski, *Surface Engineering for Wear Resistance,* Prentice-Hall, 1988

Wear of Pumps

William D. Marscher, Dresser Pump Division, Dresser Industries

Introduction

TURBOMACHINERY such as a pump is susceptible to loss of surface material from component parts by four basic mechanisms:

- Corrosion that causes pitting or the formation of surface layers of corrosion products prone to later removal
- Rubbing or sliding wear or surface fatigue in bearings and in close running clearances between stationary and rotating or sliding components
- Erosion caused by solid particles in the flow stream that scrub on flow-path surfaces or that jam in close clearances
- Erosion caused by high-frequency impact damage from liquid cavitation bubble formation and collapse in the flow stream of locations where the static pressure is locally reduced below the vapor pressure

Pitting corrosion is beyond the scope of this article. The rubbing wear, particulate erosion, and cavitation erosion mechanisms are discussed separately below. Each discussion focuses on specific components subject to the given form of wear and includes:

- The effects of wear on overall pump performance
- The relationship of the wear mechanism to relevant design parameters and environmental variables
- Specific defenses that minimize the rate of material loss, and hence extend the life of the pump, or that affect the choice of a specific pump

The pumps discussed in this article are primarily heavy-duty industrial types, including single-and multistage centrifugal, single-stage axial flow, reciprocating, and positive-displacement progressive cavity pumps, such as screw pumps.

Rubbing Wear

Centrifugal and Axial Flow Pumps. Rubbing wear in centrifugal and axial flow pumps occurs in the following components, illustrated by the multistage pump in Fig. 1:

- Impeller neck ring or wear ring annular seals, which are located in a minimum clearance area in the impeller/casing wall clearance gap to prevent leak-back of the high-pressure impeller discharge flow to the lower-pressure impeller inlet, or "suction"
- Interstage bushing or diffuser bushing annular seals, which separate the inner diameter of the backside of one stage from the inlet area of the following stage
- Axial thrust-balancing devices, such as drums with a radial annular seal or disks with an axial face seal. Such balancing devices separate multistage discharge pressure acting on one cross-sectional area from suction pressure acting on another cross-sectional area in order to oppose the accumulated axial thrust of the various pump stages. In the drum type, which is generally preferred because of its dynamic stability, if the original running clearance opens up sufficiently, parasitic leakage becomes unacceptable, and pressure values on each side change to the point that thrust-balance may no longer be possible
- Sealing between the casing fluid and the outside environment, such as "pressure-breakdown" bushing labyrinth seals, pump packing, and mechanical seals. Such seals are eliminated in magnetic-drive pumps, canned motor pumps, and most other types of hermetically sealed pumps

- Flexible couplings involving potentially sliding teeth or splines, such as gear couplings
- Bearings, including plain journals, fixed pad, tilting pad, and rolling-element bearings

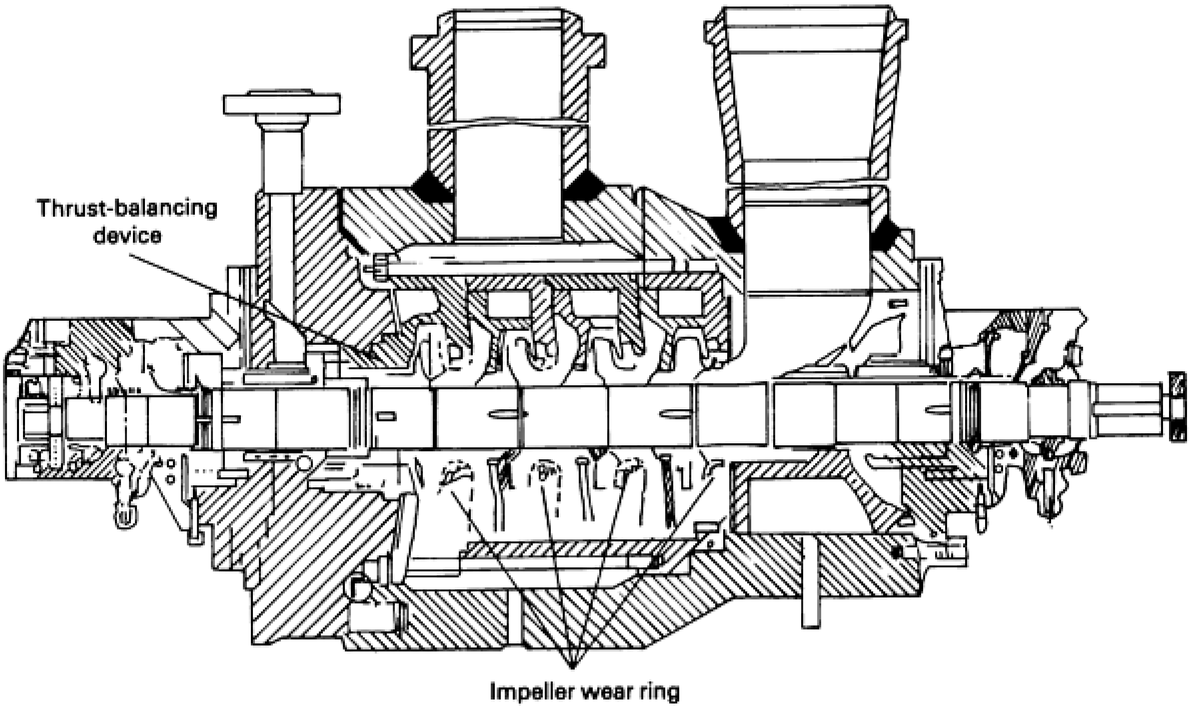


Fig. 1 Cross-sectional view of a multistage pump showing wear-prone components. Courtesy of Dresser Pump Division, Dresser Industries

These components represent the closest clearance locations within a centrifugal pump and, even in an upset condition in which one or more of them is rubbing, prevent the rotor from contacting at other locations. This same principle prevents excessive bending deflection of the rotor, such that shaft lateral steady deflection or lateral vibrations cannot induce sufficient bending stress to result in fatigue failure of the internal shaft (Ref 1). Of course, a serious metallurgical defect or machining imperfection that causes a stress concentration is always a potential failure source, whether or not there is a vibration problem.

It is common for a centrifugal pump to experience wear in one or more of the locations listed above. The goal of the pump designer and pump user should be to minimize rather than prevent rubbing wear at close clearances, so that the original clearances are at most tripled during intervals between pump overhauls. Therefore, the effects of up to three times increased clearance at all potential rubbing locations on hydraulic performance degradation must be accounted for in pump design. Generally, roughly one to two points of efficiency and 0.5 to 1% of head are lost when running clearances throughout a pump are doubled.

The predicted rotor dynamic behavior of the pump as clearances change is also critical to pump design. Multistage pump vibration can increase by a factor of 30 if the natural frequency of a shaft becomes equal to the frequency of a strong excitation force, such as residual imbalance at running speed, because this results in a resonance (Ref 2). Therefore, as a minimum, the designer must ensure that natural frequencies do not change sufficiently to cause a resonance within a range of clearance change of up to a factor of three. Otherwise, shaft vibration will exceed the running clearances, and rapid wear will result.

This can pose a significant problem for flexible shaft designs with a large number of stages, because the natural frequencies of such rotors depend strongly on the clearance of the annular seals through the workings of the "Lomakin effect" (Ref 3, 4). This effect causes a nonuniform pressure buildup through leakage velocity variation, as expressed by Bernoulli's equation. This tends to recenter the shaft if it moves off-center and therefore contributes to the supporting stiffness and thus the natural frequency of the rotor. The Lomakin in effect is illustrated in Fig. 2.

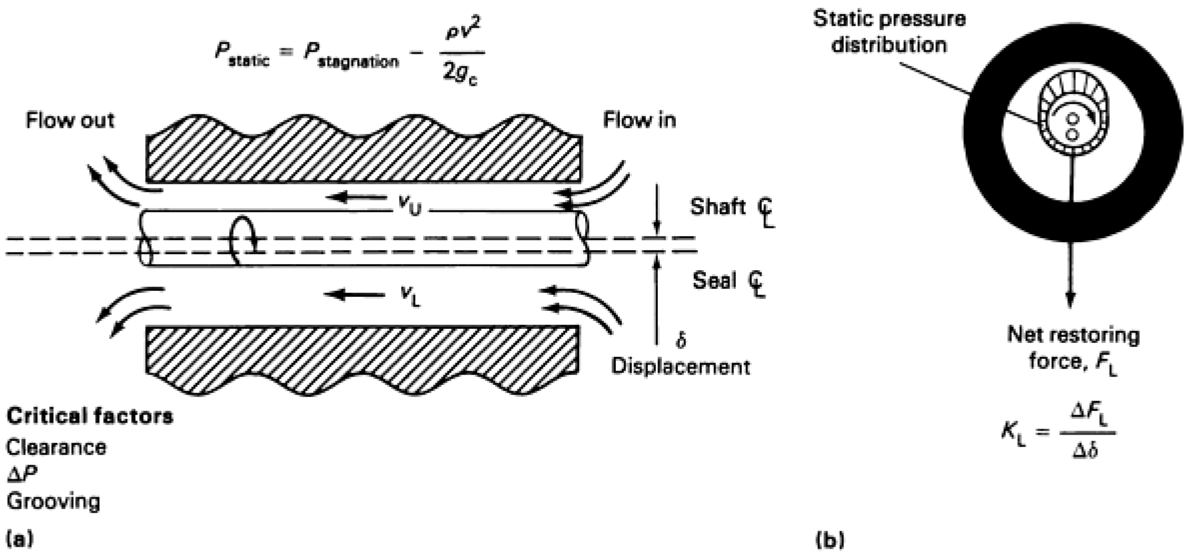


Fig. 2 Schematic showing principle of the Lomakin effect. (a) Longitudinal view of cross section. (b) End view

Even if the designer has been careful to avoid vibration-related wear problems, pump life can still be dramatically reduced unless the user (1) runs the pump within the proper flow range and (2) monitors pump vibration. The first concern is illustrated by Fig. 3, which shows a typical plot of vibration versus flow rate. Although common sense suggests that the less flow required of a pump, the less it will vibrate and wear, Fig. 3 shows that the opposite is true. The angles of the impeller and diffuser vanes of a pump can be designed to match the flow inlet and exit angles at only one particular flow rate, called the best efficiency point (BEP).

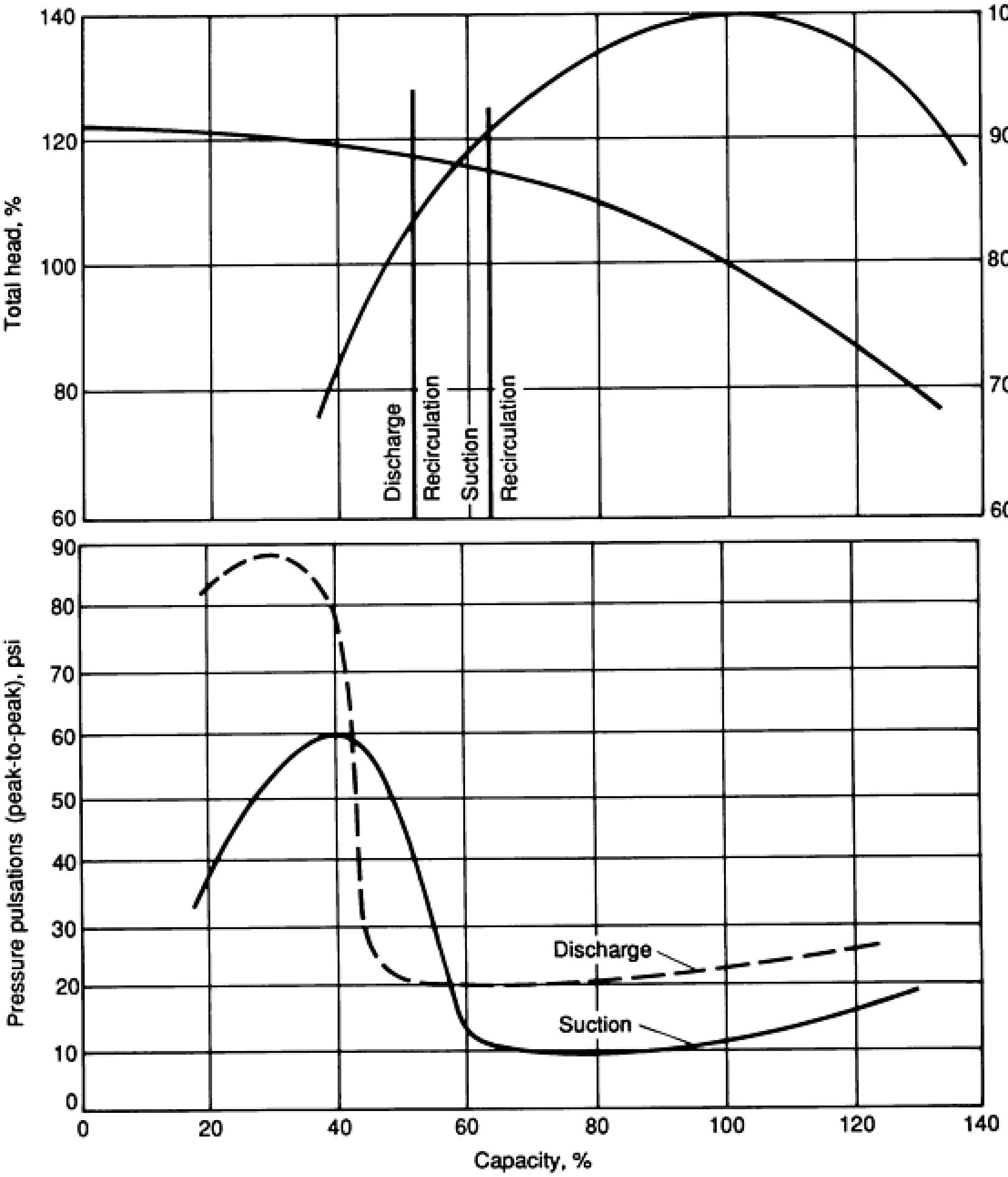


Fig. 3 Dependence of pump vibration on flow rate

The user should select a pump whose BEP is close to the expected flow requirement of a particular system, because other flow rates, higher or lower, lead to a mismatch in flow-to-vane angles and result in potentially large steady and oscillating radial loads on the pump rotor (Ref 5). If the flow mismatch becomes great enough, stalling can occur (as on an airplane wing with too high an angle of attack), leading to formation of regions of recirculating flow and an associated increase of shaking forces (Fig. 3). This is known as onset of recirculation and usually occurs first at the pump inlet or "suction" (Ref 6, 7). The minimum continuous flow of a pump should be set to avoid operating at or below this flow for extended periods of time. This sometimes requires installation of a minimum flow recirculation loop, with a valve that automatically opens when flow drops below the recirculation onset flow.

Another approach that is less wasteful than discharge throttling is to reduce pump flow as required by using a variable-frequency drive (VFD) on the pump motor to reduce operating speed. If VFDs are used, careful design of the support structures for vertically mounted pumps is essential to avoid resonances between the speed range and the structural

natural frequencies, several of which are likely to be in the speed range (Ref 8). If throttling rather than VFDs is used, other vibration problems associated with the recirculation phenomena discussed above are possible (Ref 9).

It is also important that the user monitor pump vibration levels, allowing early identification and repair of vibration problems that may accelerate wear. Various "condition monitoring" and "predictive maintenance" systems are marketed for this purpose. Figure 4 gives limits of overall vibration as a function of running speed (Ref 10). Vibration above these limits can be expected to accelerate wear at a rate such that vibration at double these limits will result in up to 10 times the rate of wear. Halving these limits generally has little or no effect, however, because at these levels no running clearance contact should have yet occurred; therefore, reducing vibration further has no impact on rubbing wear (Ref 1). Experience suggests that contact can be expected to begin, on average, at about one and a half to two times the recommended vibration limits.

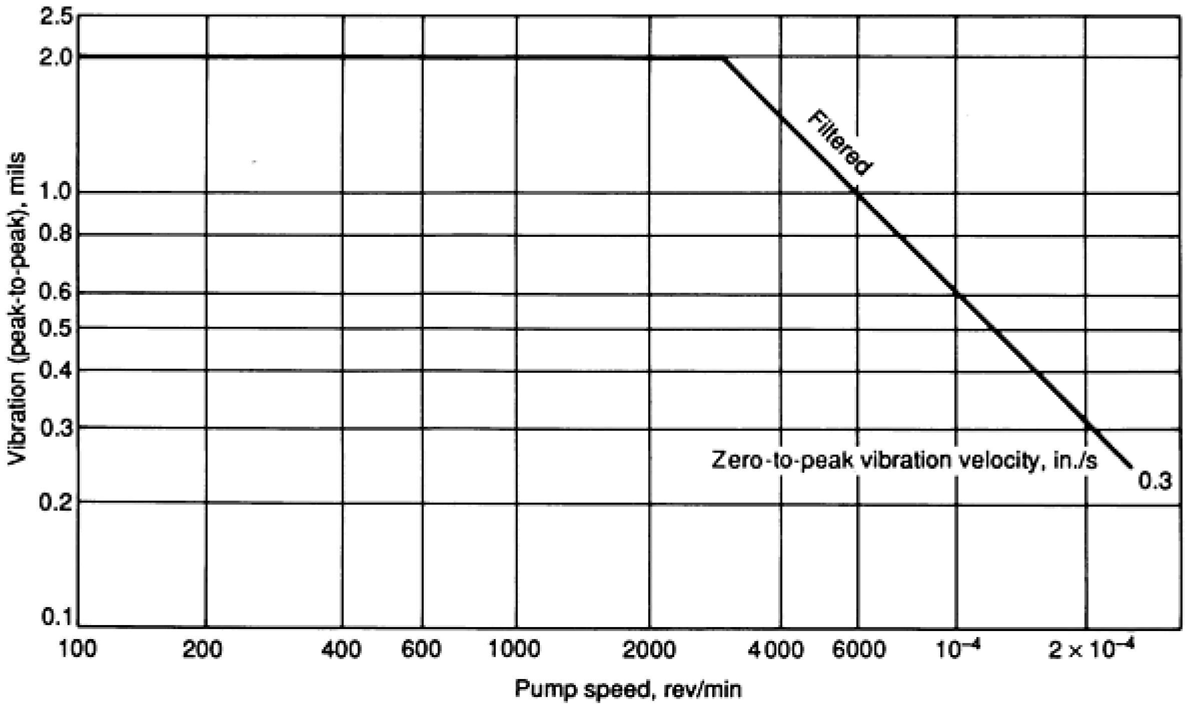


Fig. 4 Allowable vibration as a function of running speed. Adapted from Ref 10

Despite the best efforts of the designer and the user, it is likely that the occasion will arise when close clearances contact and rub. Of key importance when such contact takes place is that no galling wear occur, because this can seize the pump and cause severe damage to the rotor system or drive train by overtorque.

Galling can be resisted in several ways. Opposing material pairs can be selected from "adhesive compatibility charts," such as the one shown in Table 1, which identify those combinations with the least chemical attraction. This selection process is aided by heat treating the opposing parts so that they have a Brinell hardness difference of at least 50, which encourages wear of the softer material rather than adhesion and resultant part-to-part material transfer (Ref 11).

Table 1 Galling resistance of various material combinations

	Cast iron	3% Ni cast iron	Ni-Resist (types 1, 2)	Ductile iron	Ductile Ni-Resist	Nickel-copper alloy K-500	Nickel-copper alloy 400	Nickel-copper alloy 506	Nickel-aluminum alloy 301	Nickel 213	Nickel 305	Nickel-chromium alloy 600	Nickel-chromium alloy 705	400 series stainless steel (soft)	400 series stainless steel (hard)	300 series stainless steel	SAE 1000 to 6000 steel (soft)	SAE 1000 to 6000 steel (hard)	Bronze (leaded)	Ni-Vee bronze A	Ni-Vee bronze B	Ni-Vee bronze D	Ni-Al bronze	Hastelloy alloys A, B	Hastelloy alloy C	Hastelloy alloy D	Nitrided	Chromium plate	Stellite
Cast iron	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
3% Ni cast iron	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
Ni-Resist (types 1, 2)	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
Ductile iron	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
Ductile Ni-Resist	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
Nickel-copper alloy 505	S	S	S	S	S	F	F	S	S	S	S	F+	S	F	S	F	F	S	S	F	S	S	F	F+	S	S	S	S	S
Nickel-copper alloy K-500	S	S	S	S	S	F	N	F	F	S	S	N	S	F	F	N	N	F	S	F	S	S	F−	F	F	S	S	S	S
Nickel-copper alloy 400	S	S	S	S	S	N	N	F	N	F	F	N	S	N	F	N	N	F	S	F	S	S	F−	N	F	S	S	F	S
Nickel-copper alloy 506	S	S	S	S	S	F	F	F	F	F	F	N	S	N	F	N	N	F	S	F	S	S	F−	N	F	S	S	S	S
Nickel-aluminum alloy 301	S	S	S	S	S	F	N	F	F	S	S	N	S	N	F	N	N	F	S	F	S	S	S−	F	F	S	S	S	S
Nickel 213	S	S	S	S	S	S	F	F	S	S	F	F	S	F	S	F	F	S	S	S	S	S	S	S−	S	S	S	S	S
Nickel 305	S	S	S	S	S	F	F	F	S	F	S	F	S	F	S	F	F	S	S	S	S	S	S	S−	S	S	S	S	S
Nickel-chromium alloy 600	S	S	S	S	S	N	N	N	N	F	F	N	F	N	F	N	N	F	S	F	S	S	F−	N	F	S	S	S	S
Nickel-chromium alloy 705	S	S	S	S	S	S	S	S	S	S	S	F	S	F	S	F	F	S	S	S	S	S	S	S	S	S	S	S	S
400 series stainless steel (soft)	S	S	S	S	S	F	N	N	N	F	F	N	F	N	F	F	N	F	S	F	S	S	F−	N	F	S	F	F	S
400 series stainless steel (hard)	S	S	S	S	S	F	F	F	F	S	S	F	S	F	S	F	S	S	S	S	S	S	F+	F	S	S	S	S	S
300 series stainless steel	S	S	S	S	S	N	N	N	N	F	F	N	F	F	F	N	N	F	S	F	S	S	F−	N	F	S	S	S	S
SAE 1000 to 6000 steel (soft)	S	S	S	S	S	N	N	N	N	F	F	N	F	N	S	N	N	S	S	S	S	S	F+	N	F	S	S	S	S
SAE 1000 to 6000 steel (hard)	S	S	S	S	S	F	F	F	F	S	S	F	S	F	S	F	S	S	S	S	S	S	S	F	S	S	S	S	S
Bronze (leaded)	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
Ni-Vee bronze A	S	S	S	S	S	F	F	F	F	S	S	F	S	F	S	F	S	S	S	F	S	S	F	F	S	S	F	S	S
Ni-Vee bronze B	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	F	S	S	F	S	S	S	S	S	S
Ni-Vee bronze D	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
Ni-Al bronze	S	S	S	S	S	F−	F−	F−	F−	S	S	F−	S	F−	F+	F−	F+	S	S	F	F+	S	N	F−	F+	S	F	S	S
Hastelloy alloys A, B	S	S	S	S	S	F	N	N	F	S−	S−	N	S	N	F	N	N	F	S	F	S	S	F−	N	F	S	S	S	S
Hastelloy alloy C	S	S	S	S	S	F	F	F	F	S	S	F	S	F	S	F	F	S	S	S	S	S	F+	F	F	S	S	S	S
Hastelloy alloy D	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S
Nitrided	S	S	S	S	S	S	S	S	S	S	S	S	S	F	S	S	S	S	S	F	S	S	F	S	S	S	S	S	S
Chromium plate	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	?	S
Stellite	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S

Degree of resistance: S, satisfactory; F, fair; N, little or none. Courtesy of Goulds Pumps, Inc.

A second method of discouraging galling is to machine grooves in one or both of the close-clearance components, so that as wear takes place the debris can collect somewhere other than at the close running clearance. This also promotes rapid heat transfer at rubbing interfaces, keeping parts cool and hard. Local surface temperatures can become very high even with grooving, because of the flash-temperature effect, but such temperatures decay in short distances and do not result in

galling if surface heat removal is effective (Ref 12). Grooving the surfaces results in a design compromise, however. Although grooving reduces clearance leakage, it also reduces the beneficial shaft support provided by the Lomakin effect.

Another way to assess galling potential is with a measure number for the contact stress required for cold welding and subsequent material pullout for a material mated against itself. This is called galling threshold load and is tabulated in reference handbooks for various engineering materials. Generally, austenitic and precipitation-hardening steels have very low galling threshold loads, on the order of 15 MPa (2 ksi). On the other end of the scale is Nitronic 50, Nitronic 60, and Waukesha 88, with values of about 350 MPa (50 ksi).

In open-impeller (that is, no shroud encloses the blade passages) and axial flow pumps, such as those used in water-flood service, the impeller or propeller blades interface directly with the casing across a narrow blade-tip clearance instead of with wearing rings. The same considerations apply regarding use of compatible materials with reasonable hardness differences. For large-diameter propellers, materials should also have relatively low friction coefficients, so that sudden rubbing contact does not cause overtorque of the pump shaft or drive train. Materials selected using the compatibility chart (Table 1) generally have low friction coefficients. Typical performance loss due to blade tip wear in unshrouded impellers and propellers is about 1% efficiency and head loss for each percent of blade height removal. Open or semiopen impellers are often used for abrasive slurry service, particularly in the paper industry, as discussed below in the section on particulate erosion.

Coupling tooth or shaft spline wear is typically a reliability problem only if the teeth are not properly lubricated or if they are overloaded by a vibration resonance (Ref 13). If resonance of the shaft occurs for a rotor natural frequency that includes significant coupling motion, reciprocating slip may occur at the coupling teeth, leading to fretting wear. The only solution in this circumstance is to redesign the shaft or to use a different weight coupling to shift the natural frequency out of running speed range.

Bearing wear in pumps can take the form of gross plastic deformation, frictional softening or melting, and subsequent abrasive wear if there is a severe shaft overload or inadequate or improper lubrication. Improper lubrication in such instances is usually caused by lubricant emulsification when contaminated with water by a seal failure or by atmospheric condensation. More common than catastrophic bearing failure, however, is the gradual deterioration of the bearing by surface fatigue due to:

- Marginal lubrication, again perhaps because of lubricant contamination
- Steady loading somewhat in excess of the rated surface pressure of a journal bearing (between 1.4 and 14 MPa, or 0.2 and 2 ksi, depending on the bearing material and geometrical details) or above the rated 20,000 h B_{10} life of a rolling-element bearing
- Large oscillating loads caused by excessive vibration, usually because of either a rotor dynamic or a bearing housing natural frequency resonance (Ref 14)

In its final stages, surface fatigue is recognizable by loaded surface "craze cracking" and spalling of surface material in laminar chunks.

Vertical turbine pumps (VTPs), used for deep well, low suction head, and wet pit service, are unique with respect to their lightly loaded bearings (gravity is parallel to the shaft) and the multiple-bearing support of the often long "line-shafting" that connects the deeply submerged impeller or bowl assembly to the above-ground motor driving the pump. Another unique feature of VTPs is the "product lubrication" (that is, lubrication by the pumped fluid) of the lineshaft and bowl bearings. Bronze or bronze-filled Teflon is a good bore material in such bearings, and their wear rate has been found to be minimized by using a relatively large length-to-diameter ratio (an L/D ratio of 2.5 is common). The greater length of these bearings tends to restrict the flexure of the relatively thin lineshafting, keeping whirling amplitudes and thus bearing loads low (Ref 15). In poorly lubricating fluids, such as water, and in abrasive fluids, rubber has been found superior to bronze, particularly when combined with axial or spiral grooves to feed bearing surfaces and wash out entrapped particles. The inside surfaces of such bearings can be made polygonal or corrugated to further enhance lubricant feed and flush in the presence of low bearing loads. This style of bearing, inexpensiveand popular in irrigation pumps, gives good service as long as the pump is not allowed to run dry.

Reciprocating pumps are more prone to rubbing wear than centrifugal pumps, because they have several sliding components, including the following (Fig. 5):

- Lower crosshead versus the crosshead ways
- Tie rods versus their linear bearing surfaces
- Plunger or piston rod versus the packing
- Plunger or piston surface versus the bore walls
- Valve lateral surfaces within the valve retainer or guide

In addition, each suction and discharge valve is subject to impact wear at the seating surface, particularly if cavitation is occurring. Finally, there are several sets of highly loaded bearings: crankshaft bearings (one for each one or two cylinders plus the coupling), connecting rod "small-end" bearings attached to each lower crosshead pin, and connecting rod "large-end" bearings that attach each rod to the crankshaft.

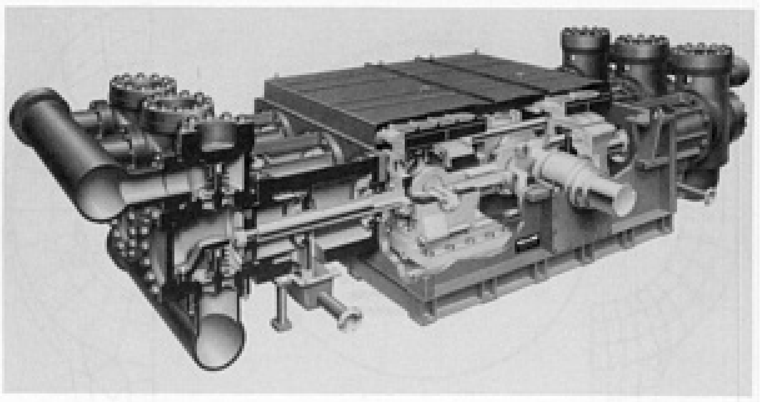


Fig. 5 Cutaway view of a reciprocating pump. Courtesy of Dresser Pump Division, Dresser Industries

To minimize wear in reciprocating sliding components, it is critical to provide clean flushing fluid to unlubricated areas and to filter the oil and rigorously maintain its quality for lubricated areas. Ceramic plungers and bore liners together with clean flushing fluid are effective for avoiding excessive wear induced by third bodies during abrasive service. In the crosshead ways, excellent alignment must be maintained with the crosshead to avoid locally high bearing pressures, and area-averaged surface bearing pressure during the highest load portion of the stroke should be kept well under 690 kPa (100 psi) by properly sizing the crosshead bearing surface during design.

Valve surfaces and their retainers must be maintained at a high hardness (on the order of 40 HRC), with a hardness differential of 50 HB, and must be properly selected from the compatibility chart (Table 1). In addition, valves should be designed and tested such that absolutely no cavitation occurs during the initial valve opening, when static pressure is at its lowest. Otherwise, severe valve impact or "hammer" can occur, damaging the valve and the seat and, through the transmitted elastic wave, jeopardizing the entire drive train (Ref 16).

Finally, special vertically mounted spherical or conical seat valves should be used in slurry or abrasive pumping service or if deposits are expected to build up near the cylinders. Such valves tend to discourage the trapping of debris in the seating interface (Fig. 6) (Ref 17).

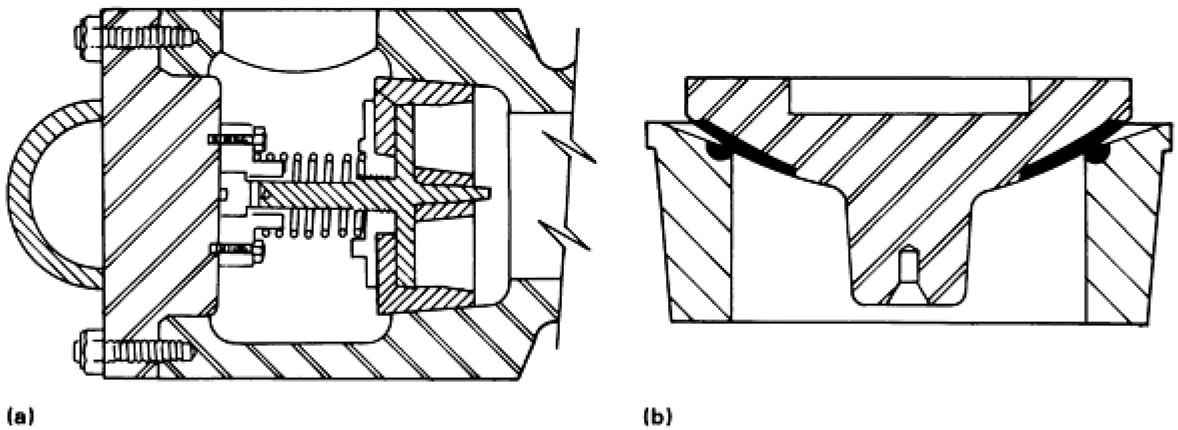


Fig. 6 Cross-sectional view of abrasive-resistant slurry valves used in reciprocating pumps. (a) Rib-supported conical type. (b) Spherical type. Source: Ref 17

Reciprocating pump journal bearings tend to be highly loaded, with surface pressures from 6.9 to 13.8 MPa (1 to 2 ksi). Special bearing babbitt materials are needed, and configurations adapted to reciprocating motion, such as "bedded arc" nonsymmetrical bore designs for connecting rod small-end bearings, are used for higher load capacity (Ref 18).

It is also important that the bearing bore and journal be kept aligned and that the journal bore clearances be maintained. This requires careful assembly, as well as a design procedure that accounts for transient thermal differences in the bore versus the journal and for elastic deformation of the bearing supports. For example, if connecting rod "eye" cross sections at either the small end or large end are not sufficient, the bearing can ovalize to the extent that all of its side clearance is taken up, "pinching" the journal and causing possible seizure of the bearing (Fig. 7).

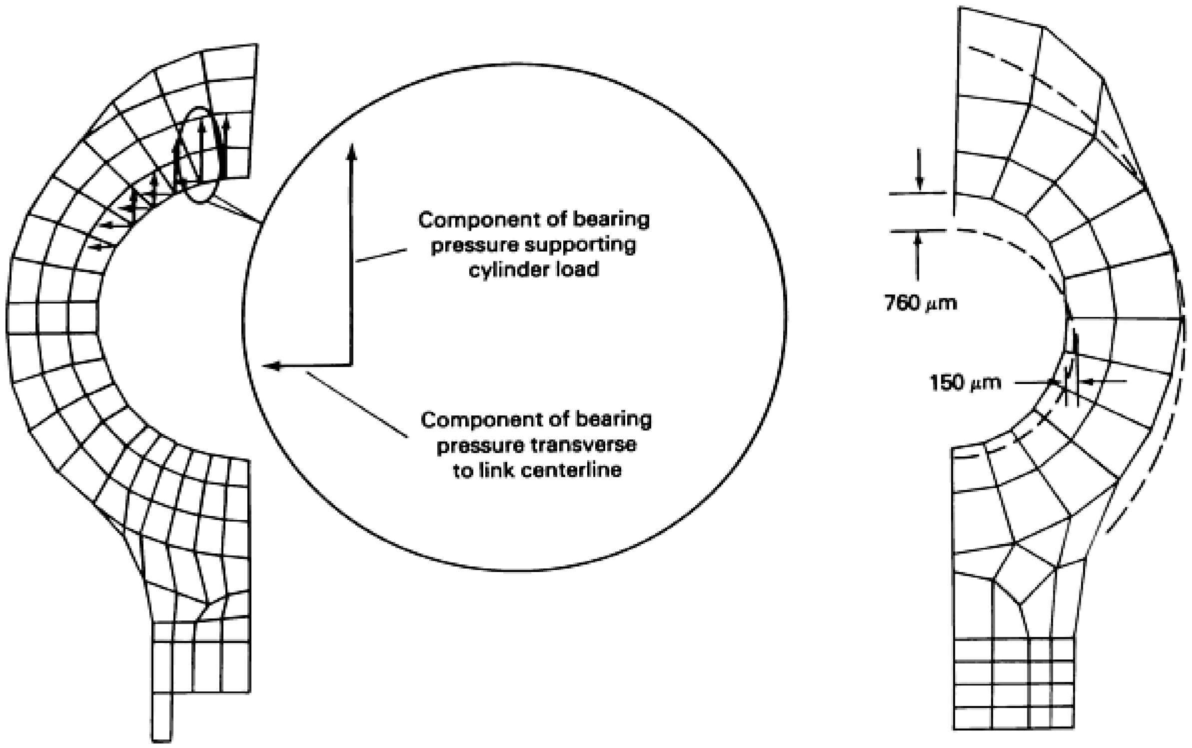


Fig. 7 Bearing "pinch" generated by deformation of a small-end connecting rod

The crankshaft bearing is also prone to elastic support problems. If the crankshaft is relatively stiff, care must be taken that the bearing support stiffness of the pump frame is optimized to carry the load and yet allow the bearings to translate and thus distribute the plunger loads of the rigid crankshaft to several bearings. If the crankshaft is relatively flexible, on the other hand, the bearing support stiffness must be adjusted to avoid locally high bearing pressure at bearing corners when the crankshaft tilts within the bearing shell.

Particulate Erosion

Problems of third-body erosion can occur in all types of pumps if they are subjected to abrasive service, such as slurry service. Commonly pumped slurries include coal, mined phosphate, mined metal ore, and water with fly ash captured during flue-gas desulfurization. The pumping of paper pulp is another common erosive service. Such services must compromise hydraulic performance for wear resistance and require designs that allow ease of inspection and repair (Ref 19).

The abrasivity of a fluid can be quantified using the Miller number index, as outlined in ASTM Standard G 75 (Ref 20). The effective viscosity of an abrasive slurry is often quite high compared with the virgin fluid and should be considered along with abrasivity in selecting an appropriate pump. Otherwise, the pump may have a head and flow capability that are too low for the intended service.

Centrifugal Pumps. The typical centrifugal slurry pump is of an overhung single-stage end-suction design. It should have reasonably low required value of net positive suction head (NPSH), as explained below in the section on cavitation. This is because suction piping velocities must often be kept high so that the slurry stays above its "critical carrying velocity" or "nonsettling velocity" below which the slurry cannot be kept evenly suspended (Ref 21). This requirement is particularly strong if the slurry particles are larger than roughly 65 μm.

The areas of a centrifugal pump most prone to abrasive wear are the impellers and the casing interior, where flow-path velocities are high. One approach to decreasing erosive wear of these areas and to maintaining the pump is the use of hard, wear-resistant, replaceable liners. Elastomeric liners are sometimes used as an alternative to hard metal liners. Natural rubber, neoprene, nitrile, and chlorosulfonated polyethylene have exhibited high abrasion resistance when used as external liners for impellers and as internal liners for casing volute passages. However, care must be taken that the pressure in the flow passage next to the liner does not fall below the pressure between the liner and the wall, which is typically close to atmospheric pressure. When this occurs in service, unless the liner is mechanically bolted to the casing at key locations, it may billow out, catching the impeller blades and destroying the rotor. Another caveat is to stay below the softening temperature of the rubber, typically about 115 °C (240 °F). Also, for elastomeric materials to exhibit good abrasion resistance, the abrasive particles must be less than 5 mm (0.2 in.) if they are dull and less than 2 mm (0.08 in.) if they are sharp. Finally, the head per stage should be kept below about 50 m (165 ft) (Ref 19).

When metal impellers and casing liners are used, they should be as hard as possible, not only because wear resistance generally increases with hardness, but also because abrasive erosion in pump components generally occurs due to the impingement of particles at very oblique angles to the surface. As pointed out by Finnie (Ref 22), such abrasion attacks ductile materials much more readily than it does hard, relatively brittle materials. Finnie's equation for hard materials is:

$$\text{Wear rate} = (\text{Number of impinging particles}) \cdot (\text{Average particle mass}) \cdot (\text{Impingement velocity})^2 \cdot (\text{Angle of impingement}) \qquad \textbf{(Eq 1)}$$

Particulate erosion of a ductile steel impeller by a catalytic fluid that was not expected to be erosive is shown in Fig. 8.

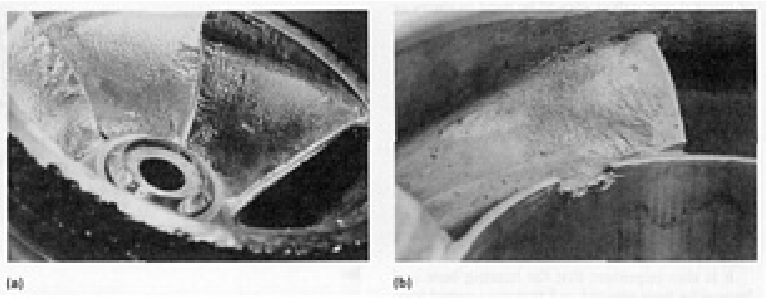


Fig. 8 Particulate erosion of a ductile steel impeller by an abrasive catalyst. (a) 0.25×. (b) 1×

Typical hard, erosion-resistant metals include white cast iron (standard gray or ductile cast irons are poor), high-chromium (that is, 13 to 28%) alloy steel, cobalt-base superalloys such as Stellite, and special nickel-base alloys such as Ni-Hard. These materials are useful not only for the flow-path surfaces but also as sleeves in sealing areas, particularly if packing is used. It is good practice to use hard metal or ceramic expellers near seal entrances to keep particles out of the sealing clearance. If thorough flushing of packing presents a design problem, abrasion-resistant SiC ceramic mechanical seals together with an expeller or minimal flushing can be used. Some of the newest designs of mechanical seals for abrasive service are even axially split, for ease of replacement.

Design steps other than material selection can be taken to maximize the ability of a pump to withstand abrasives. For example, the casing volute tongue can be retracted to avoid the highest-velocity region of discharge flow, and casing volute walls can be contoured to minimize impingement angles in high-density zones of particle trajectories (Ref 23). The impeller can be redesigned with its blades recessed into toroidal cavities, thus avoiding many mainstream particles (however, a stiff efficiency penalty can be expected for this configuration). Impeller vanes can be made thicker so that it takes longer for their edges to erode. Also, larger pumps than required can be selected and then run at less than design speed, through use of a simple adjustable pulley or variable-frequency drive excitation of the motor (Ref 24).

Even if optimum materials and design configurations are chosen, centrifugal pumps still face practical limitations in abrasive service. For example, head per stage should be kept below 100 m (330 ft) or else the implied volute velocities will be too high. When higher heads are required, pumps can be arranged in series. Also, abrasive particles may collect in impeller side clearances, held in suspension by the counter-tendencies of the leakage flow, which is biased to moving toward lower diameter by design, and of centrifugal forces, which always work away from minimum diameter. This can result in long-lived, particle-filled vortices (Ref 25), which quickly erode the impeller hub unless hardened wear plates are applied at the proper locations.

Open or semiopen impellers are often used for abrasive or slurry service, particularly in the paper industry. Such impellers can be designed with clearance in the axial direction, which sacrifices new pump efficiency but makes the pump performance less prone to clearance increases and allows clearance adjustment by shimming of the impeller to change its position on the shaft. The surfaces of the impellers and casing flow paths, particularly the volute, must be made of a hard metal or ceramic or of a tough elastomer, as discussed above.

Reciprocating and Other Positive-Displacement Pumps. Generally, centrifugal pumps are the best choice if relatively low heads and high flows are required; otherwise, reciprocating pumps are the best choice. Other criteria, however, are involved in abrasive service. As discussed above, centrifugal pump heads face certain limitations in such applications. On the other hand, for particle sizes greater than about 2.5 mm (0.1 in.), even the most particle-tolerant reciprocating pump valves will tend to jam and are not satisfactory.

In such cases, a diaphragm pump driven by air or clean fluid may be suitable up to about 100 m (330 ft) of developed head. Above this pressure, a progressive cavity pump might be considered (Ref 26). One type of progressive cavity pump, the single- or double-screw pump shown in Fig. 9, is capable of high pressure and relatively high flow. In abrasive service, screw pumps have been shown to pump abrasives and even mixed-phase fluids reliably with the use of hard metal screws and high-durometer (about Shore 68) rubber liners.

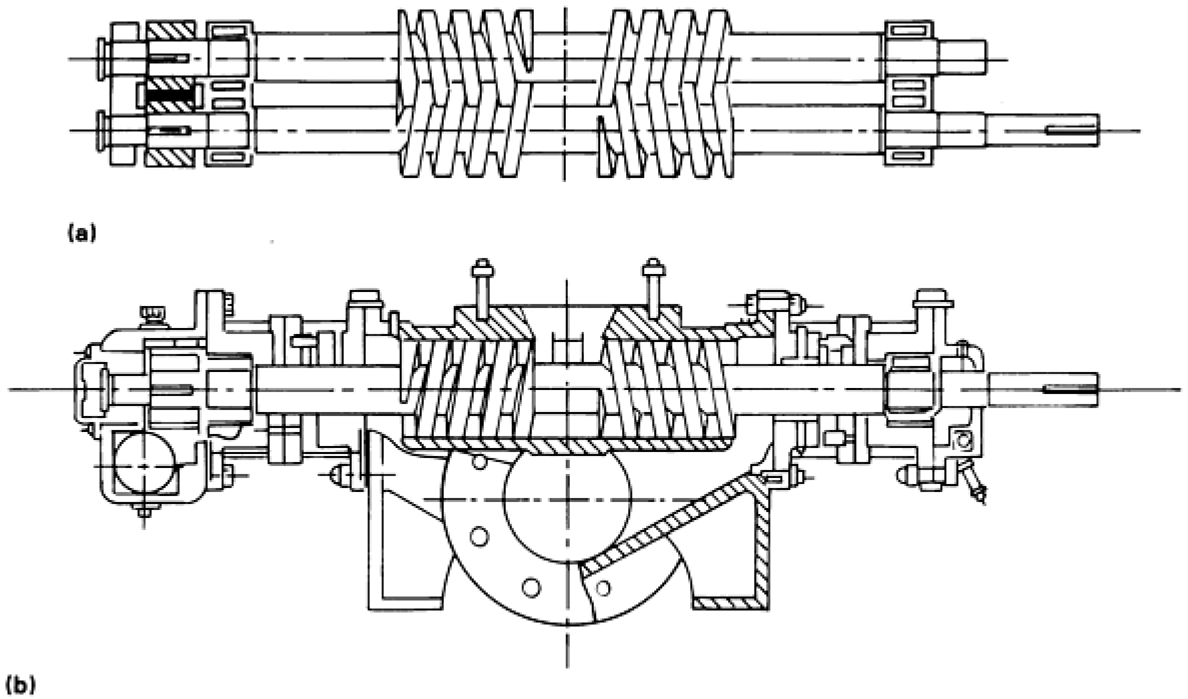


Fig. 9 Cross-sectional view of a double-screw pump. (a) Top view. (b) Side view. Courtesy of Dresser Pump Division, Dresser Industries

If particle size is below 2.5 mm (0.1 in.), reciprocating pumps are applicable, and in services where they compete with centrifugal pumps, they may even be the economical choice because of their high thermodynamic efficiency (on the order of 95%) (Ref 27). Two classes of reciprocating pumps are piston and plunger types. The piston types are generally superior for higher flows, but the plunger types can better tolerate sharp and highly abrasive particles.

Special vertically mounted spherical seat or reinforced conical seat valves should be used in slurry or abrasive pumping service, or if deposits are expected to build up near the cylinders. Such valves tend to discourage the trapping of debris in the seating surface, as illustrated by the valves shown in the earlier discussion of rubbing wear (Fig. 6) (Ref 17).

Cavitation Erosion

Cavitation occurs in portions of the flow path where static pressure drops below the vapor pressure of the pumped fluid. This causes very small vapor bubbles to form, which then rapidly and violently collapse when they enter an area where static pressure once again exceeds vapor pressure. The sudden bubble collapse causes concentrated shock waves that impinge on nearby metal, resulting in stresses large enough to induce fatigue.

Unlike surface fatigue caused by rubbing, which produces subsurface cracks that eventually delaminate the surface in thin flakes, many cavitation fatigue cracks form perpendicular to the surface and tend to remove material in small blocks. The resulting damage leaves a surface finish that initially is matte, later becomes pock marked, and eventually has the appearance of rough, mountainous terrain, with odd eroded surface protrusions. Cavitation is probably the most common cause of rapid component erosion and wear in hydraulic machinery.

The usual measure of likelihood of cavitation is the net positive suction head (NPSH), which is the difference between the static pressure and the vapor pressure of the liquid at the pump inlet flange, expressed in meters:

$$\mathrm{NPSH} = \frac{P_s - P_v}{\rho g} \quad \textbf{(Eq 2)}$$

where P_s is the static pressure at the pump inlet flange (in Pa), P_v is the vapor pressure of the liquid (in Pa), ρ is the liquid density (in kg/m^3), and g is the gravitational constant, 9.8 m/s^2. A required value of this number, the NPSHR, is determined by the pump manufacturer, generally on the basis of when sufficient bubble formation occurs to drop the head produced by the pump by 3%. The system's available NPSH, or NPSHA, must be at least equal to this value. If it is not, either another pump that is less demanding at the suction must be chosen or some means must be found to increase the NPSHA, such as an upstream booster pump or, in a centrifugal pump, use of an axial inducer. The latter looks like a corkscrew protruding from the first-stage inlet and controls the inlet flow angle while adding to the initial suction pressure. Care must be taken to include in NPSHA the losses due to suction pipe friction, valves, and strainers. In reciprocating pumps, the effect of pulsating flow inertia, or "acceleration," can reduce transient suction pressure dramatically and must be accounted for and reduced, if necessary, by using a pulsation dampener.

The situation became complicated in recent years when researchers found that NPSHR based on 3% head drop was insufficient to avoid cavitation damage. Instances occurred in which more than three times as much suction head was required to avoid bubble formation and subsequent damage. Other more stringent criteria have been proposed, such as "1%" and "0%" (that is, on the verge of total head drop) (see Ref 28), but the pump industry still generally uses the familiar 3%, with a factor of safety based on experience for a given class of pumps. Some researchers believe that observation of cavitation beyond a certain bubble length should be the criterion (Ref 29), but so far, general application of this has been impractical.

An early attempt to predict cavitation used the Thoma number, which is a ratio of NPSHA to total head. This number, originally applied to hydraulic turbines, places too much emphasis on discharge conditions to be generally applicable to pumps. A modification is the cavitation number or coefficient, *K*, which is a ratio of NPSHA to suction tip speed squared. *K* is referenced to the suction, but does not account for local increases in velocity and associated decreases in static pressure in the flow passages, especially in the turbulent and sometimes partially stalled inlet of impellers.

Another factor of importance is the ratio of NPSHR--NPSHA to vapor pressure, where NPSHR is the value of NPSH at which cavitation damage begins (not necessarily the same as the 3% NPSHR, as discussed above). Once cavitation damage begins, this ratio is higher for cold fluids than for warm ones, because a warm fluid has a higher vapor pressure than a cold one. This higher ratio suggests that cavitation is more severe in cold fluids, which is exactly the case. A warm fluid, with higher vapor pressure, will cavitate at a lower suction pressure; however, the rate of bubble collapse and hence the severity of the cavitation once it occurs is roughly dependent on the ratio of the static pressure differential to the vapor pressure, so the bubble collapse damage is less severe.

Centrifugal Pumps. A useful approach for assessing the potential for cavitation damage in centrifugal pumps is the concept of suction-specific speed, *S*, developed by Karassik (Ref 30):

$$S = \frac{[\text{rpm} \cdot (\text{flow})^{1/2}]}{(\text{NPSHR})^{3/4}} \qquad \textbf{(Eq 3)}$$

As *S* becomes larger, the chances become greater that cavitation damage will occur despite the 3% head drop NPSHR value being met, and if it does occur it is more severe, on average, as *S* increases. Cavitation in pumps without inducers becomes highly probable as *S* increases beyond the range of 13,300 to 16,000, where NPSH is in feet and flow is in gallons per minute (Ref 31).

An indirect but important cause of cavitation is suction and discharge recirculation (Ref 6, 7), which occurs when the pump is run well below the flow it was designed for, and the inlet or exit flow angles match poorly with the pump vane angles. This leads to local secondary flow eddies on the high-pressure side of the blade. The locally increased velocity in these eddies can decrease local suction pressure sufficiently to cause cavitation and subsequent damage. This damage is usually distinguishable from direct cavitation, because it usually occurs on the pressure side of the blade. Direct cavitation generally occurs where the flow-path static pressure is low in the primary flow--namely, on the suction surface of the blades, usually close to the inlet. An example of direct cavitation damage is shown in Fig. 10.

Fig. 10 Early stages of centrifugal pump cavitation and damage

As discussed earlier in the section on rubbing wear, recirculation and its associated problems can be avoided by ensuring that the continuous (that is, long-term) flow of the system is not below the flow associated with the onset of suction recirculation, which is typically somewhere between 25 and 75% of the rated capacity of the pump. This may be difficult in an existing installation, and a redesigned impeller to lower recirculation onset flow should be considered, although this may reduce efficiency somewhat.

In instances of either direct or recirculation-induced cavitation damage, if it is not possible to directly fix the cause (that is, by increased suction pressure or increased minimum flow, respectively), then a material with greater cavitation erosion resistance can be used to replace the existing impellers. In general, cast iron, low-carbon steel, and austenitic steels have low cavitation resistance, although they are often used where significant cavitation is not expected. Titanium, aluminum bronze, Nitronic 50, and a new alloy called Hydrolloy have good cavitation erosion resistance, the tradeoff being increased expense.

A standardized accelerated test procedure (ASTM G 32) (Ref 32) is available for assessing the cavitation damage resistance of materials if existing data cannot be found. This procedure is based on the use of high-frequency vibrations to excite cavitation close to the metal surface.

Reciprocating Pumps. The components of a reciprocating pump most likely to suffer cavitation damage are the suction valves. Suction valves should be designed and tested for a given pump flow and quoted NPSHR such that absolutely no cavitation occurs during the initial valve opening, when the annular valve area is small and fluid velocity is high, causing locally low pressures through Bernoulli's equation. When even a small amount of cavitation occurs in one valve, it triggers an "impulse response" of the suction manifold acoustic natural frequencies, with resulting pressure pulsations large enough to make all of the suction valves severely cavitate. This can cause "valve hammer," where the valve sharply impacts its seat at an acceleration in excess of 100 G's. The resulting shock tends to peen the valve seating area, causing fatigue, and can send strong stress waves through the entire drive train. This can loosen bolted connections and cause fatigue or bearing problems far from the source of the problem (Ref 16).

If suction valve cavitation occurs, either suction pressure must be increased or the valve design must be changed to lower the NPSHR. When determining reciprocating pump suction pressure, it is important to include the suction line inertia effects when the suction valve opens, known as acceleration head. This reduces the effective NPSHA and can be minimized by shortening the suction piping or by placing a large-volume pulsation dampener in the piping, as close to the pump as possible (Ref 17).

Additional information about cavitation damage can be found in Ref 33 and in the article "Cavitation Erosion" in this Volume.

References

1. W. Marscher, "The Relationship Between Pump Rotor System Tribology and Appropriate Vibration Specifications for Centrifugal Pumps," Paper C123/87, Institution of Mechanical Engineers, 1987
2. W. Marscher, Structural Design and Analysis, *Sawyer's Gas Turbine Handbook,* D. Japikse, Ed., Turbomachinery International Publications, 1985

3. H. Black and D.N. Jenssen, Dynamic Hybrid Properties of Annular Pressure Seals, Proc. J. Mech. Eng., Vol 184, 1970, p. 92-100
4. W. Marscher, Analysis and Test of Multi-stage Pump "Wet" Critical Speeds, *Tribology Trans.,* Vol 34 (No. 3), 1991, p 445-457
5. A. Agostinelli *et al.,* An Experimental Investigation of Radial Thrust in Centrifugal Pumps, *J. Eng. Power (Trans. ASME),* Vol 82 (No. 2), 1960
6. W. Fraser, "Recirculation in Centrifugal Pumps," presented at ASME Winter Annual Meeting (Washington, DC), Nov 1981, p 1-22
7. B. Schiavello and M. Sen, On the Prediction of the Reverse Flow Onset at the Centrifugal Pump Inlet, *Proceedings of the ASME Symposium on Performance Prediction of Centrifugal Pumps and Compressors,* American Society of Mechanical Engineers, 1980
8. W. Marscher, Reliability of Vertical Pumps, *Proceedings of Water and Wastewater Conference '90* (Barcelona), 1990
9. W. Marscher, "Subsynchronous Vibration in Boiler Feed Pumps Due to Stable Response to Hydraulic Forces at Part Load," Paper C349/88, Institution of Mechanical Engineers, 1988
10. API-610 Pump Specification, 7th ed., American Institute of Petroleum Engineers, 1989
11. W. Marscher, a Phenomenological Model of Abradable Wear in High Performance Turbomachinery, *Wear,* Vol 59, 1980, p 191-211
12. W. Marscher, "A Critical Evaluation of the Flash Temperature Concept," ASLE Preprint 81-AM-1D-3, American Society of Lubrication Engineers, 1981
13. A. Crease, Design Principles and Lubrication of Gear Couplings, *Proceedings of IMechE International Conference on Flexible Couplings,* Institution of Mechanical Engineers, 1976
14. W. Marscher, Vibration Test and Analysis of a Barrel Boiler Feed Pump Exhibiting Non-synchronous Vibration, *Proceedings of IMechE Seminar on Vibration in Centrifugal Pumps,* Institution of Mechanical Engineers, 1990
15. W. Marscher, The Effect of Fluid Forces at Various Operation Conditions on the Vibrations of Vertical Turbine Pumps, *Proceedings of IMechE Seminar on Pumping,* Institution of Mechanical Engineers, 1986
16. S. Collier, Know Your Triplex Mud Pump, Parts 1-6, *World Oil Mag.,* Jan-June 1982
17. J. Miller, "Reciprocating Pumps for Slurry Service," ASLE Preprint 84-AM-6A-1, American Society of Lubrication Engineers, 1984
18. J. Campbell *et al.,* Bearings for Reciprocating Machinery: A Review of the Present State of Theoretical, Experimental, and Service Knowledge, *Proceedings of the IMechE Conference on Lubrication and Wear,* Institution of Mechanical Engineers, 1968
19. J. Doolin *et al.,* Pumping Difficult Fluids, *Chem. Eng.,* Dec 1991, p 67-79
20. "Test Method for Slurry Abrasivity by Miller Number," G 75, *Annual Book of ASTM Standards,* Vol 3.02, ASTM, 1984, p 420-437
21. G.R. Addie, "Centrifugal Slurry Pump Selection and Application Tutorial," presented at 4th International Pump Symposium, Texas A&M University, 1987
22. I. Finnie, Some Observations on the Erosion of Ductile Metals, *Wear,* Vol 19, 1972, p 81-90
23. M.C. Roco and G.R. Addie, Analytical Model and Experimental Studies on Slurry Flow and Erosion Flow and Erosion in Pump Casings, *Proceedings of the Slurry Transfer Association,* March 1981
24. W. O'Keefe, Pumps, Valves, and Piping, *Power Mag.,* March 1992, p 19-30
25. A.P. Smith, Unexplained Wear in Large Centrifugal Pumps, *Chem. Eng.,* Aug 1979, p 153-155
26. A. Prang, "Rotary Screw Pumps for Multiphase Products," presented at Conference on Seals and Vibration Reliability of Centrifugal Turbomachinery (Ukraine), 17-20 Sept 1991
27. W. Smith, Construction of Solids Handling Displacement Pumps, *Power Fluids Mag.,* Vol 9 (No. 1), 1983
28. E. Grist, "Net Positive Suction Head Requirements for Avoidance of Unacceptable Cavitation Erosion in Centrifugal Pumps," Paper C163/74, Institution of Mechanical Engineers, 1974

29. J. Gulich and A. Rosch, Cavitation-Erosion in Centrifugal Pumps, *Sulzer Tech. Rev.,* No. 1, 1988, p 28-32
30. I. Karassik, Centrifugal Pumps and System Hydraulics, *Chem. Eng.,* Oct 1982, p 84-106
31. J. Hallam, Centrifugal Pump Suction Specific Speed, How Important Is It?, *Proceedings of the ASME Petroleum Mechanical Engineering Conference,* American Society of Mechanical Engineers, 1982
32. "Method for Vibratory Cavitation Erosion Test," G 32, *Annual Book of ASTM Standards,* Vol 3.02, ASTM, 1988
33. J. Doolin, Judge Relative Cavitation Peril with Aid of These Eight Factors, *Power Mag.,* Oct 1985

Friction and Wear of Compressors

Royce N. Brown, Dow Chemical Company

Introduction

A COMPRESSOR is a mechanical device that is used to increase the pressure level of compressible fluid. These pressures can be at any level, from vacuum to high positive levels. The fluids can consist of either vapors or gases, singularly or as mixtures. Because the compressors described in this article use mechanical motion to perform the compression function, they are subject to friction and wear. Each of these compressors is defined below by type and function, followed by a brief discussion of how friction and wear are manifest in each.

Reciprocating Compressors

The reciprocating compressor is one of the more complex types of machines, in terms of moving parts. It performs compression by trapping a volume of gas in a cylinder that contains a piston. The piston displaces and compresses the gas on the head end stroke (Fig. 1). The two most common forms of reciprocating compressor are the trunk and crosshead types.

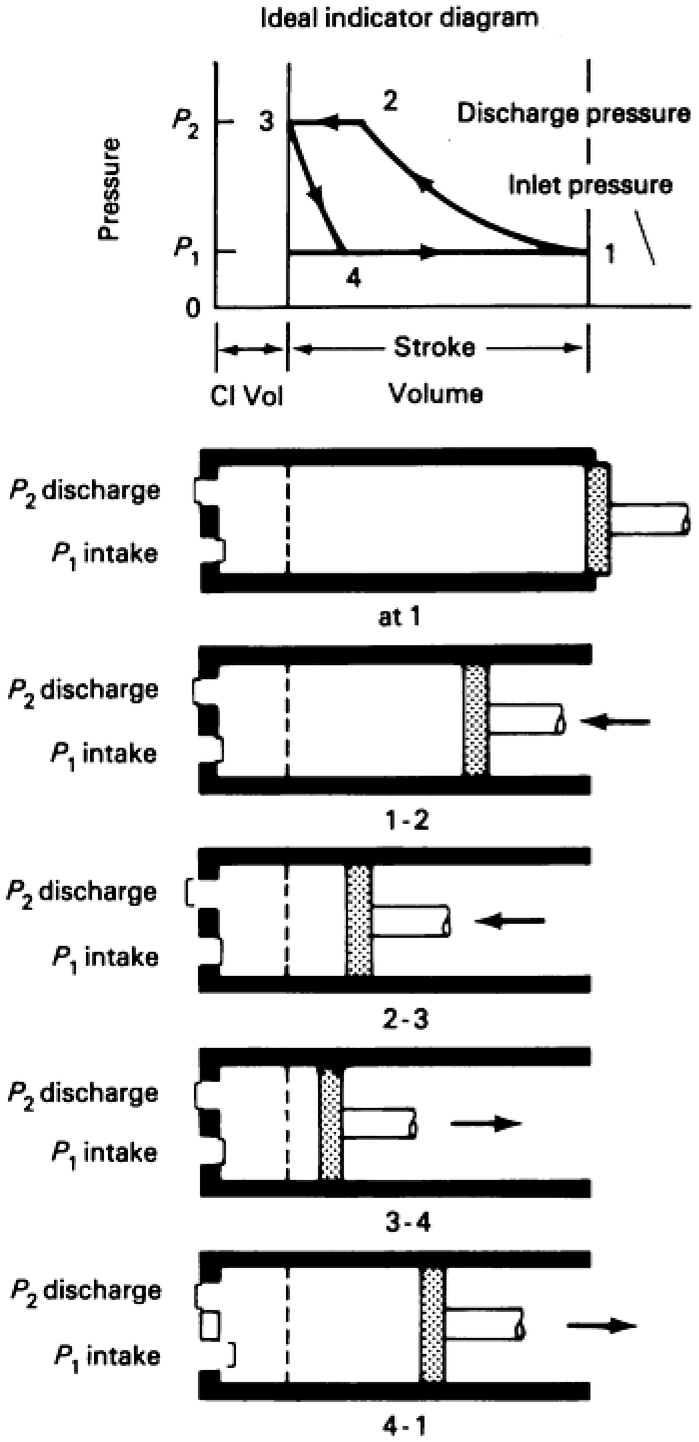


Fig. 1 Reciprocating compressor cycle

The distinguishing feature of the trunk-type compressor (Fig. 2) is that the piston is directly connected to the connecting rod. This type of construction is normally used in smaller machines, such as small shop air compressors and refrigeration service compressors.

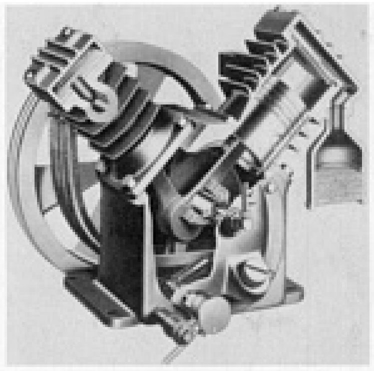


Fig. 2 Trunk-type reciprocating compressor. Courtesy of Ingersol Rand

The crosshead-type compressor (Fig. 3) is normally found in process gas service. The connecting rod is attached to the crosshead, which is connected to the piston rod that moves the piston. The cylinders are normally double acting, a feature that permits compression to occur on either side of the piston, which essentially doubles the capacity for a given-sized cylinder. It can be noted that a gas seal in the form of packing can prevent gas leakage from the compressor cylinder to the crankcase.

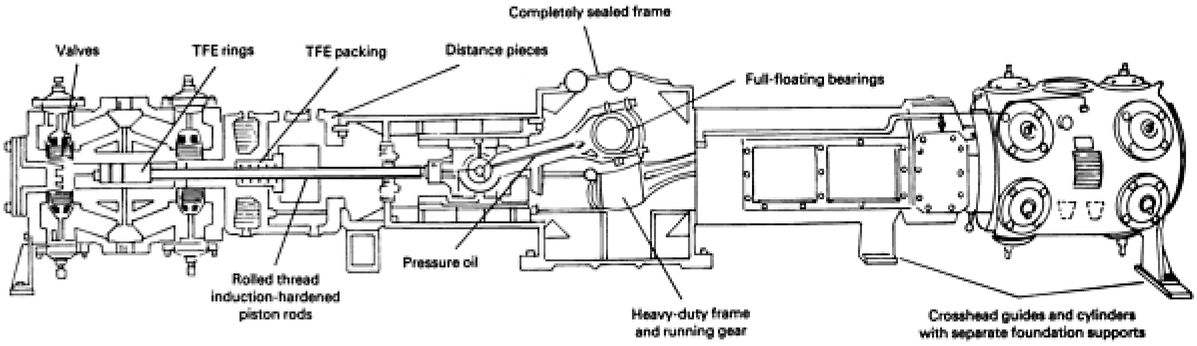


Fig. 3 Cross section of crosshead-type reciprocating compressor. Source: Dresser-Rand

The cylinder, which is the gas-handling part of the reciprocating compressor, contains the piston, which is the gas displacer. Motion is transmitted to the piston by the connecting rod in the trunk-type compressor, and by the piston rod in the crosshead type. The cylinder includes the valves that control the inlet and outlet gas flow. The compressor valves are spring loaded and gas actuated. There are one or more inlet valves and one or more discharge valves. Flow and pressure considerations would dictate the number and when multiple valves are needed.

Cylinder Wear. The piston includes a set of seals, more commonly referred to as piston rings. Because the rings rub the cylinder walls, wear of both the rings and the cylinder wall becomes a consideration in material selection. The majority of cylinders used for process gas service utilize replaceable liners that act as the cylinder wear surface.

Cylinder orientation has an influence on the wear of the cylinder and rings. Vertically oriented cylinders would be preferred if wear were the only consideration. Small compressors are generally constructed with the cylinders vertical, or nearly so in a "V" configuration. As compressor size increases, height also increases, which causes head-room problems and makes maintenance unwieldy.

Generally, the larger compressors are arranged with horizontal cylinders; this arrangement makes spaces management easier and facilitates maintenance. However, these gains are somewhat offset by the effects of gravity. The weight of the piston causes additional rubbing forces to the cylinder wall and increases the potential for wear, particularly to the rings. To counter this, nonmetallic (usually) rider rings are supplied with the pistons. The rider rings are larger and act as the

wear surface for the piston. American Petroleum Institute (API) Standard 618 provides guidelines on the loading limits of rider rings used in reciprocating compressors for refinery service.

As stated earlier, the piston rings act as a seal to minimize the quantity of gas that can bypass the piston. Again, the need for compromise becomes evident. If there are many rings and/or a very tight fit to the cylinder wall, then sealing is excellent, but friction force and wear are increased. Besides the wear considerations, increased friction results in a larger drag on the piston. This absorbs energy and results in an increased parasitic load on the compressor. Ultimately, this loss results in a decrease in efficiency. A piston is shown in Fig. 4.

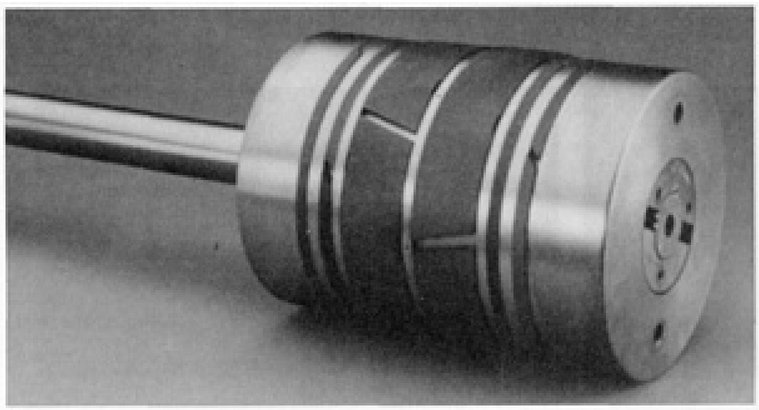


Fig. 4 Crosshead-type reciprocating compressor piston and rod. Courtesy of Dresser-Rand

Piston speed is a key factor in the wear of a cylinder and the rings. Average piston speed is defined as:

$$PS = 2SN \qquad \textbf{(Eq 1)}$$

where PS is piston speed in m/s, S is stroke in m, and N is crankshaft speed in rev/s. Piston speeds normally range from 2 to 6 m/s (400 to 1200 ft/min). It is obvious that the higher the piston speed, the higher the wear. Industrial compressors are normally selected in the range from 3 to 4 m/s (600 to 800 ft/min). If the cylinder is of a nonlubricated design, then the piston speed should be toward the lower end of the given speed range to minimize wear.

Cylinder lubrication depends on the type of compressor being considered. In trunk-type compressors, cylinder lubrication is normally obtained from the crankcase, and reaches the cylinder wall by a splash action. Although not sophisticated, this method does provide adequate lubrication.

In crosshead-type compressors, cylinder lubrication is divided into either nonlubricated or lubricated cylinder compressors. The nonlubricated cylinder compressor does not supply an outside source of lubricant to the cylinder. The term nonlubricated is probably not the most accurate description of the lubrication process. In actuality, the ring materials are chosen for their inherent low friction or self-lubricating characteristics. The lubricated cylinder compressor is furnished, by an outside source, with a lubricant that is either a conventional mineral oil or a synthetic oil, depending on the process gas.

Two methods of supplying the lubricant to the cylinder and rod packing are available. The amount of lubricant furnished to any lubrication point is quite small, on the order of droplets per minute. Use of the metering-type lubricator permits precise tailoring of lubricant quantity to the cylinder points and packing without compromising the frame lubrication system, which is much larger.

One type of mechanical lubricator consists of a multiplunger pump, where an individual plunger is dedicated to each lubrication point. A sight glass is normally included in each plunger feed to allow visual monitoring of the point. Another common type of mechanical lubricator system consists of a single metering pump sized for the total combined flow to all points. A divider block distributes the flow in the required amount to the individual lubrication points. Each system has

pros and cons associated with it. Generally, the compressor supplier can provide recommendations to users who have not developed a preference.

Cylinder Materials. In many smaller compressors, the wearing portion of the cylinder is made of the same material as the balance of the cylinder (usually, gray cast iron). In some instances, a cylinder can be nodular, or ductile, iron. On smaller high-pressure service cylinders, steel is used. Smaller compressors of the more standardized type normally use an unlined cylinder.

On larger compressors, cylinders with replaceable liners are more common. Liner material can be chosen for its wear properties, as well as compatibility with the application. If cylinder wear does occur, then the liner construction represents a renewable part. For an unlined cylinder, the only recourse after significant cylinder wear is to either rebore or replace the cylinder.

The piston ring materials are made of either a fluorocarbon compound or polyether etherketone (PEEK) for nonlubricated service. The fluorocarbon is generally a carbon-filled type. For lubricated service, rings can be either metallic or nonmetallic. Bronze and cast iron are common for metallic service, whereas carbon-filled fluorocarbon is normally used for nonmetallic service. Rider rings generally use carbon-filled fluorocarbon.

Valves are used to control the gas that either enters or leaves a cylinder. Valves can be one of the higher-maintenance items on the reciprocating compressor. Because valve duty is quite severe, it is subject to cyclic stresses that are due to the requirement that it open and close on each stroke. Upon each open and close cycle, the valve is subjected to impact loads. It also experiences all of the temperature changes that occur in the compressed gas. On lubricated cylinders, the presence of lubricant does not extend valve life. Valve motion is difficult to lubricate. Also, in some cases, the lubricant in combination with high temperatures may cause carbon to form on the valves, shortening their lives. As the valves wear, compressor performance deteriorates, which lowers efficiency and raises temperature, furthering the deterioration problem.

A plate valve is shown in Fig. 5 and a channel valve is shown in Fig. 6. In most cases, the material of construction is metallic. Gray iron, nodular iron, and steel are commonly used for valve seats and guards. Valve springs are chrome, silicon, stainless steel, or inconel. Higher alloys may be needed for reactive process gases. Valve plates are usually stainless steel, plastics, or PEEK.

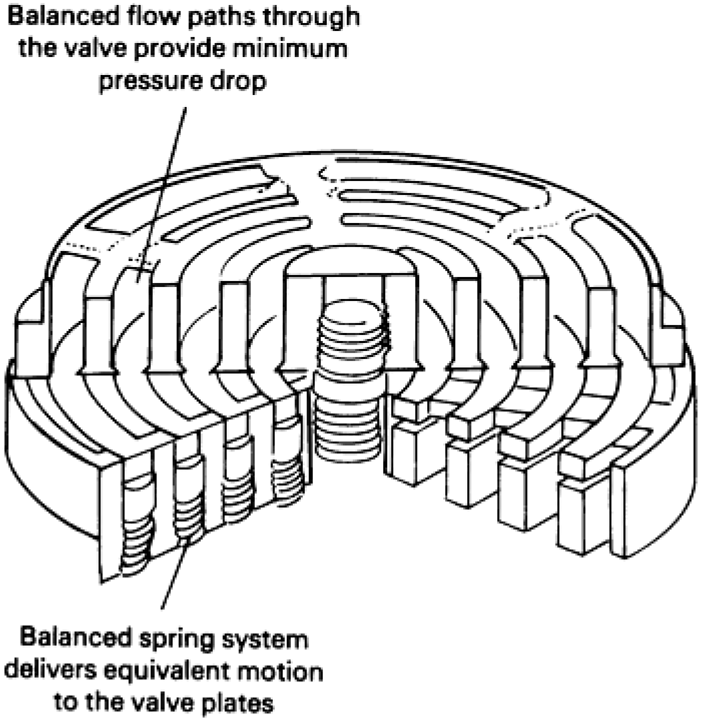


Fig. 5 Plate valve. Source: Dresser-Rand

Fig. 6 Channel valve. Courtesy of Dresser-Rand

Piston rod packing is required on double-acting cylinders where the piston rod passes through the crank-end cylinder closure. This may also be true on the head end if a tail rod is used. The purpose of the packing is to contain the gas in the cylinder. The piston rod must have a high degree of finish in the packing area. Various methods of surface hardening are used. These features are needed to minimize the packing and rod wear in the packing region. On lubricated compressors, the packing is given lubrication from the cylinder lubricator, which helps reduce the friction and wear at this point. The packing rings are generally a segmental type, and are usually made from a carbon-filled fluorocarbon-type material.

In addition to being a source of wear, packing is another source of friction, which can contribute to loss in efficiency. Gas release that is due to excessive leakage is an extreme result of excess wear.

The crosshead is the point where the connecting rod motion is converted to pure linear motion. It provides the feature that makes double-acting cylinders possible. A crosshead and associated parts are shown in Fig. 7. The crosshead is lubricated by the frame lubrication system. The critical requirement for the crosshead is that the piston rod provide a force reversal at the crosshead pin. This reversal causes the pin to change its position in the crosshead bearing clearance. This movement allows oil to enter both sides of the bushing to ensure adequate lubrication.

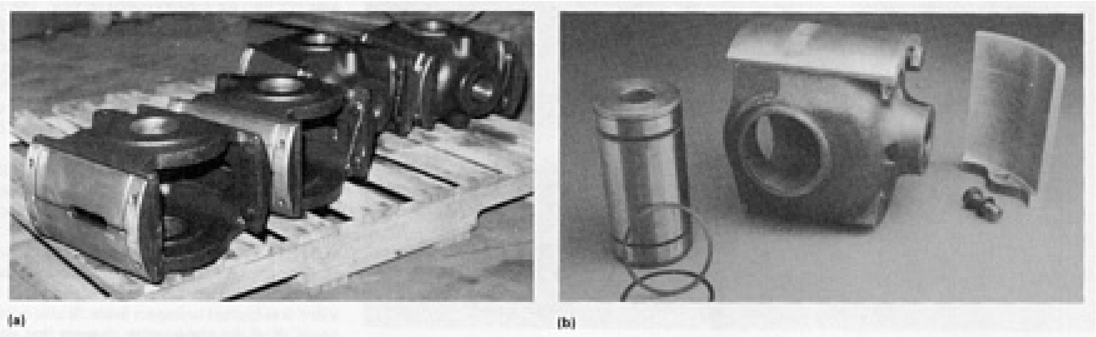


Fig. 7 (a) Crossheads. (b) Crosshead with pin and shoes. Courtesy of Dresser-Rand

On a fully loaded cylinder, rod reversals are not a problem. However, if one end of the double-acting cylinder is unloaded, the gas load could exceed the inertial load and prevent the reversals from taking place. Clearance pocket unloaders or external clearance bottlers can be used for alternate unloading schemes to solve rod reversal problems.

The frame portion of the reciprocating compressor is a major component of the compressor, in terms of physical size and weight. It houses the crankshaft, where rotary motion is translated into the reciprocating motion required to move the pistons. It provides the mounting base for the cylinders and provides for their alignment to the crankshaft, which is cradled in a set of main bearings located in the frame. The crankshaft also has journals for the connecting rod bearings.

These bearings can be either the hydrodynamic oil film type or the rolling-element type. There may also be a combination; the main bearings could be one type and the connecting rod journals, the other type. Normally, the larger process compressors use the hydrodynamic type for all bearings.

On the crosshead-type compressor, the crosshead housing is attached to the basic frame member, and the distance piece and cylinder, in turn, attach to the crosshead housing. In some designs, the crosshead housing is integral to the basic frame member.

It could be said that the frame makes a contribution to valve life. This connection is not really all that mysterious, in that the crankshaft speed dictates the valve cycle rate. If the crankshaft speed is high for a given service, then the valve cycle rate is high. The cycle rate has a direct impact on valve life.

Frame Lubrication. The frame section of the reciprocating compressor is independently lubricated in crosshead-type machines. For the trunk type, it has been noted that all lubrication, both for frame and cylinders, is supplied by the frame system. One method used in smaller machines is a very simple nonpressurized system, which lubricates by splashing the oil. The splashing is accomplished by the dipping of the crank throws into oil contained in a sump in the crankcase. Dippers can be attached to the connecting rods to aid in the oil distribution.

Pressurized systems have at least one oil pump, which is driven by either the compressor's crankshaft or a separate electric motor. There can be additional pumps, such as a start-up pump for shaft-driven main pump configurations. The auxiliary pump may alternately take the form of a full-sized electric motor standby. A filter that provides for oil cleanup can be duplicated to allow servicing while the compressor is operating.

When the heat duty cannot be dissipated in the sump, an external oil cooler is required. This exchanger is either a liquid-cooled or air-cooled device. For critical service, two coolers can be used to permit servicing while the unit is operating. Service valves and instrumentation for monitoring pressure, temperature, and level also may be included depending on the critical nature of the service and on the size and type of unit.

Reciprocating Compressor Overview. The interaction of economics and life is an ever-present challenge to the user of any equipment, but this is particularly true for the reciprocating compressor. If, for example, a volumetric rate of gas is specified, then several choices are available. First, a cylinder size and speed are selected, and then a small-diameter long stroke can be chosen. This has some first-cost advantage over choosing a larger bore, shorter stroke. The disadvantage is that the piston speed may be too high for good life. Crankshaft speed is also a factor. A higher speed will reduce the overall size of the cylinder and again lower the first cost, but this impacts both piston speed and valve cycle rate.

The final selection is normally the best compromise of first-cost and compressor component life. This will vary from application to application, based on the critical nature of the service and the ability to withstand downtime. Even if downtime is not considered serious, another factor is the projected cost of maintenance and its impact on the life cost of the machine.

Rotary Compressors

Many rotary compressor types have a low-pressure design and are generally referred to as blowers. Their pressure differential capacity is normally limited. This discussion focuses on rotary compressors with a somewhat higher pressure range, as well as the capability for a higher pressure rise. To some degree, these compressors are capable to competing with the reciprocating compressor.

Like the reciprocating compressor, the rotary compressors are of the positive-displacement type. In actuality, they supplement the range of applications of the reciprocating compressor, rather than duplicate it.

Lobe Type

Lobe-type compressors all have some features in common. However, the most significant type is the helical lobe, or screw compressor, which is also referred to as the SRM and Lysholm compressor. It can be configured as either a non-flooded (dry) or flooded compressor. The dry configuration is generally larger in capacity, but is limited to a lower pressure ratio. The flooded compressor is used for lower volumetric flow, but can achieve a higher pressure ratio.

The dry helical lobe compressor consists of two rotors, one male and one female, which are arranged in their casing to intermesh when they turn. The compression cycle is shown in Fig. 8. The gas path takes the form of the helix cut in the female rotor. The rotors are timed by a set of timing gears located at one end of the machine. Gas path sealing is accomplished by seal strips located at the tips of the rotors. Two rotor profiles are shown in Fig. 9.

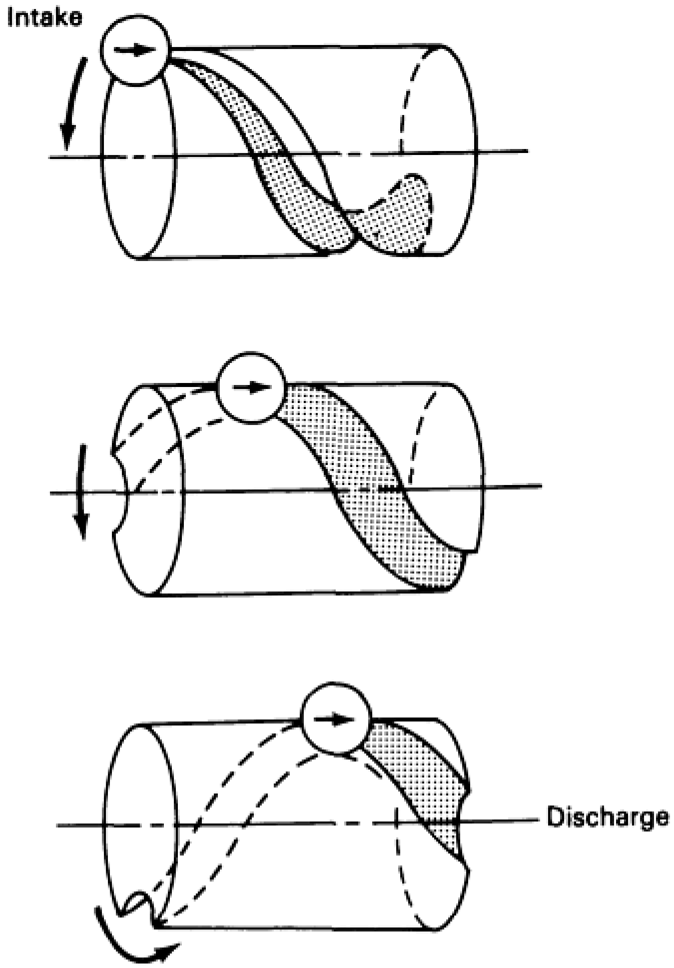


Fig. 8 Compression cycle of helical lobe compressor

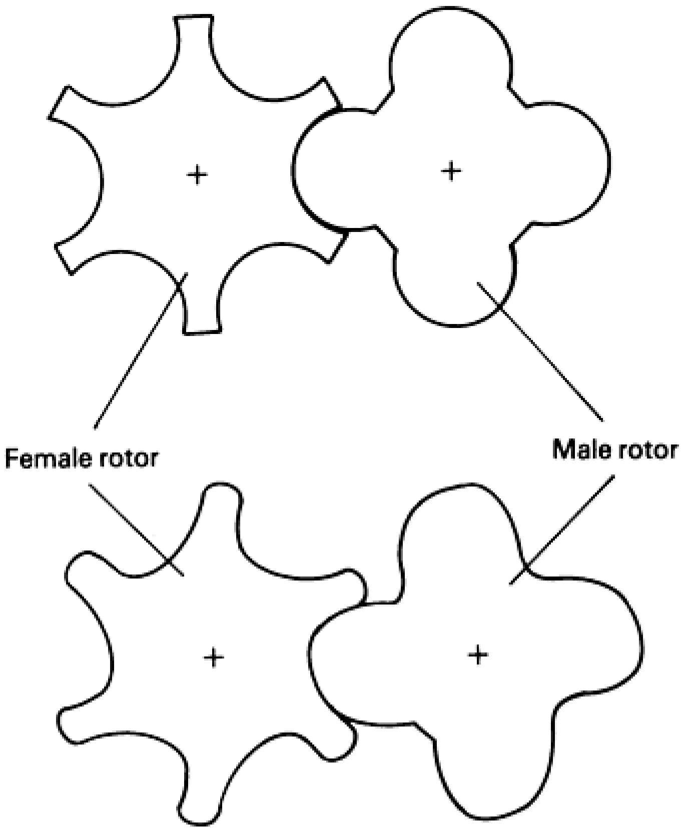


Fig. 9 Rotor profiles

The rotor discharge ports are cut to match the application pressure ratio. The purpose is to release the pocket of trapped gas being compressed at the application level to prevent overcompression and undercompression. This allows the compressor to operate at its optimum efficiency. A set of helical screw compressor rotors is shown in Fig. 10.

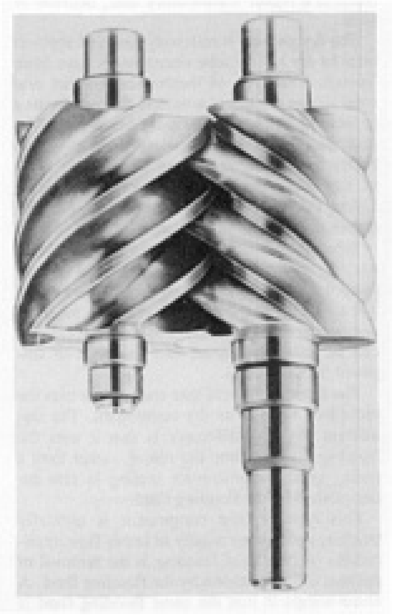


Fig. 10 Helical screw compressor rotors

Rotor Wear. The rotors, in ideal service, would not wear because they would not touch, being synchronized by the timing gears. Realistically, wear does occur. Although this style of compressor works better in dirty gas service than many other compressors, the presence of particulates are what causes wear. Normally, this occurs on the seal strips. This compressor also tolerates liquid mist better than the other compressors. If this type of operation is anticipated, the supplier can overlay the casing with a welded material to add abrasion or corrosion resistance and to minimize rotor bore wear. Although seal strips can be renewed, it is difficult and therefore expensive to repair a badly eroded casing.

Bearings and Seals. Dry helical lobe compressors normally have hydrodynamic radial and thrust bearings, which are pressure fed from an outside oil system. Because of the nature of the oil film that develops from rotation, hydrodynamic bearings do not suffer from much wear. Wear normally is due to the lack of lubricant cleanliness. If the particle size in the lubricant is larger than the minimum lubricant film thickness, then excess wear occurs. This is true for the four journal bearings and the thrust bearing. A well-maintained lubrication system with properly sized oil filters should prevent this.

There are a few applications for which rolling-element bearings may be used by the manufacturer. However, this is not very common on dry helical lobe compressors used in process service.

The dry helical lobe compressor requires four shaft end seals, which can have many configurations. The application dictates the basic requirements of leakage. On an air compressor, a relatively simple labyrinth or carbon ring seal is adequate. This type of seal is generally referred to as a restrictive seal, because it limits, rather than stops, leakage. Of the two, the carbon ring seal is more effective, but unfortunately is more subject to wear than the plain labyrinth.

Two other basic forms of seals are commonly used, depending on application requirements: the liquid film and the mechanical contact seal. The liquid film seal uses a metallic sealing ring, which is liquid buffered to maintain a liquid film in the ring clearance area. The presence of the liquid film excludes the gas, thereby effecting a seal. The liquid normally is a lubricant, which helps prevent wear to the bushing. Water can be used for applications where the lubricant and gas are incompatible. Carbon bushings are sometimes used in water-buffered service.

The mechanical contact seal is more expensive, but also provides the greater degree of gas containment. This seal type can have either a liquid buffered or dry gas form. The preferred fluid for buffering is a lubricant. When lubricating oils is compatible with the process, lubricant from the bearing lubrication system can be used. The control of the lubrication system pressure becomes more complicated, because several pressure levels may be required. However, no additional liquid supply system is needed. In some rare applications, this seal is used with a liquid that has poor lubrication properties. The result is a higher maintenance cost, because of wear.

The dry gas seal is relatively new and applications in dry helical lobe compressors have been limited. This type of mechanical contact seal uses a clean dry gas as a buffer. The gas forms a separating film so that the seal does not rub. The seal uses less-parasitic power than its liquid buffered counterpart. It can be obtained in a variety of configurations. To prevent premature failure that is due to excessive wear, the seal buffer gas must be well filtered.

When the application requires very low leakage, the mechanical contact seal should be considered. However, as of early 1992, no commercial seal with zero leakage is available for these compressors. With careful engineering, leakage can be either contained and retrieved or disposed, in most cases.

The flooded helical lobe compressor uses the same lobe form as its dry counterpart. The significant physical difference is that it uses the flooding fluid to time the rotors, rather than a timing gear. Rotor-to-rotor sealing is also accomplished by the flooding fluid.

This type of lobe compressor is normally smaller, and is used mostly in lower-flow applications. A benefit of flooding is the removal of the heat of compression by the flooding fluid. A disadvantage is that the same flooding fluid is used to lubricate the bearings. If the gas is not compatible with lubricating oil, then special bearings may be required to operate with a gas-compatible flooding fluid.

Because of the size of the flooded compressor, rolling-element bearings are commonly used, although hydrodynamic bearings are also quite common. Seals are of the buffered mechanical contact type, because the sealing fluid must be contained with the gas.

An advantage of the flooded helical lobe compressor is that it is a relatively simple machine. It can accommodate a moderately high pressure ratio in a single casing. Therefore, it is more compact than some of its competitors. Its major

disadvantage is due to the same feature that provides its advantage: the flooding fluid. Incompatibility with the fluid limits the possible applications. The fluid must be removed from the gas stream to prevent interference with the process. This requires complex separation equipment in some applications. Additionally, the auxiliary system to supply the lubrication and flooding fluid is quite large. In many cases, it dwarfs the compressor.

Sliding Vane

The sliding vane compressor is a positive-displacement rotary-type compressor. It consists of a slotted rotor mounted eccentrically in a cylinder that is slightly larger than the rotor. The rotor slots hold a set of vanes, which are free to move radially within the slots. Centrifugal force moves the vanes out of their slots and keeps them in contact with the cylinder wall. A cross section of this compressor is shown in Fig. 11.

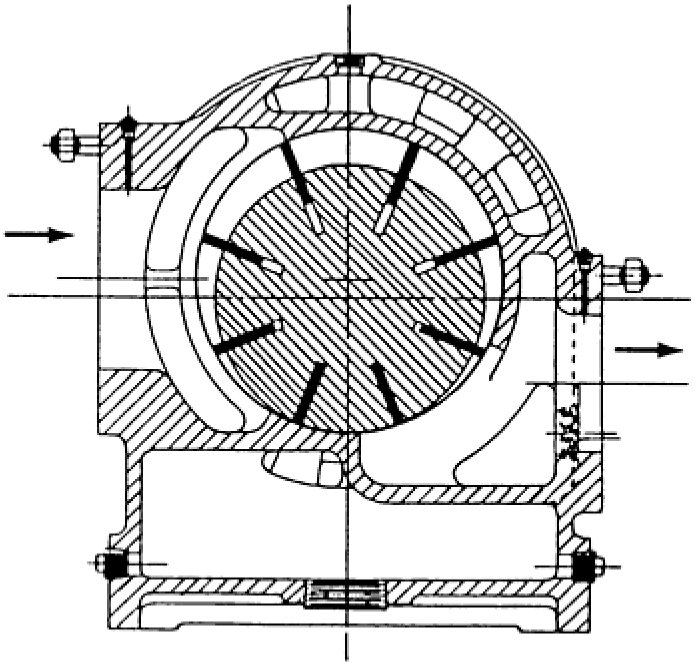


Fig. 11 Cross section of sliding vane compressor. Courtesy of AC Compressor Corporation.

The space between the rotor, set of vanes, and the cylinder forms the gas cell. The inlet is located such that the cell expands when passing the inlet port. The cell then decreases in size as the rotor continues to turn, compressing the gas. The outlet port is strategically located to be passed when the compression matches the outlet conditions. This compression process is similar to the one described for the lobe compressor.

Components. The cylinder is most commonly made of cast iron, with the bore being machined to a fine finish. The vanes are made of a laminated construction impregnated with a phenolic resin. The vanes are in continuous rubbing contact with the cylinder wall, which causes vane wear. This, of course, makes vane replacement a routine maintenance item.

To minimize the effects of vane friction and wear, most of the rotary vane compressors inject a lubricant into the cylinder. A mechanical plunger-type metering pump is commonly used, and is similar to the point lubricator described for the lubricated reciprocating compressor. Although there is much less lubricant than would be used in a flooded helical lobe compressor, it represents a disadvantage if there are gas or process compatibility problems. Its removal further compounds the problem.

Rolling-element bearings are commonly used in all but the largest of these compressors. With the plunger-per-point lubricator, a point is normally devoted to each bearing, which takes care of bearing lubrication.

Centrifugal Compressors

The centrifugal compressor normally is applied in applications where the volumetric flow is larger than that which is typical of positive-displacement compressors. Compared to the latter, the pressure ratio across a typical centrifugal compressor is less.

This compressor is different from the positive displacement type because it causes gas to be compressed by using gas dynamics. The active element of the centrifugal compressor is the impeller. All work done to the gas occurs in the impeller. The balance of the gas path components are used to convert the energy in the gas from velocity head to pressure. Other components are used to guide the gas through the compressor with minimum losses.

The impeller consists of a set of vanes enclosed in a front and back shroud, as in an enclosed design. The front shroud can be stationary, with the vanes attached to the back shroud, as in a semienclosed design. Sometimes, the vanes are simply supported, in spider form, as in an open design.

As the impeller rotates, a quantity of gas is taken up into the passage between the vanes. The shape of the impeller vanes and rotation speed move the gas along the passage, imparting energy to the gas. Some of the energy is converted into the velocity of the gas. Depending on the design, the balance of the energy conversion is to pressure. As with all compressors, a temperature rise is associated with the compression.

Types. Centrifugal compressors are available in many different configurations, the most basic of which is the overhung single-stage compressor, that is, a single impeller in an overhung configuration (Fig. 12). One method of compounding the centrifugal compressor is to use the overhung single-stage compressor with a gear. Gas passes from one stage to another on the gear frame by way of coolers between the stages. This is a common configuration for plant air, and is shown in cross section in Fig. 13. An assembled gear-mounted centrifugal compressor is depicted in Fig. 14. A form of multistaging is shown in Fig. 15. Figure 16 shows a straight-through multistage centrifugal compressor built with a horizontally split casing.

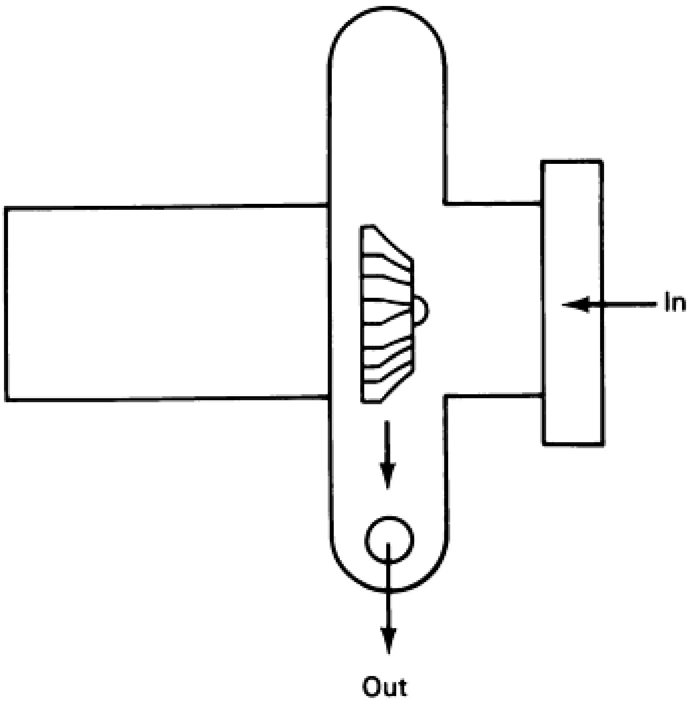


Fig. 12 Single-stage compressor

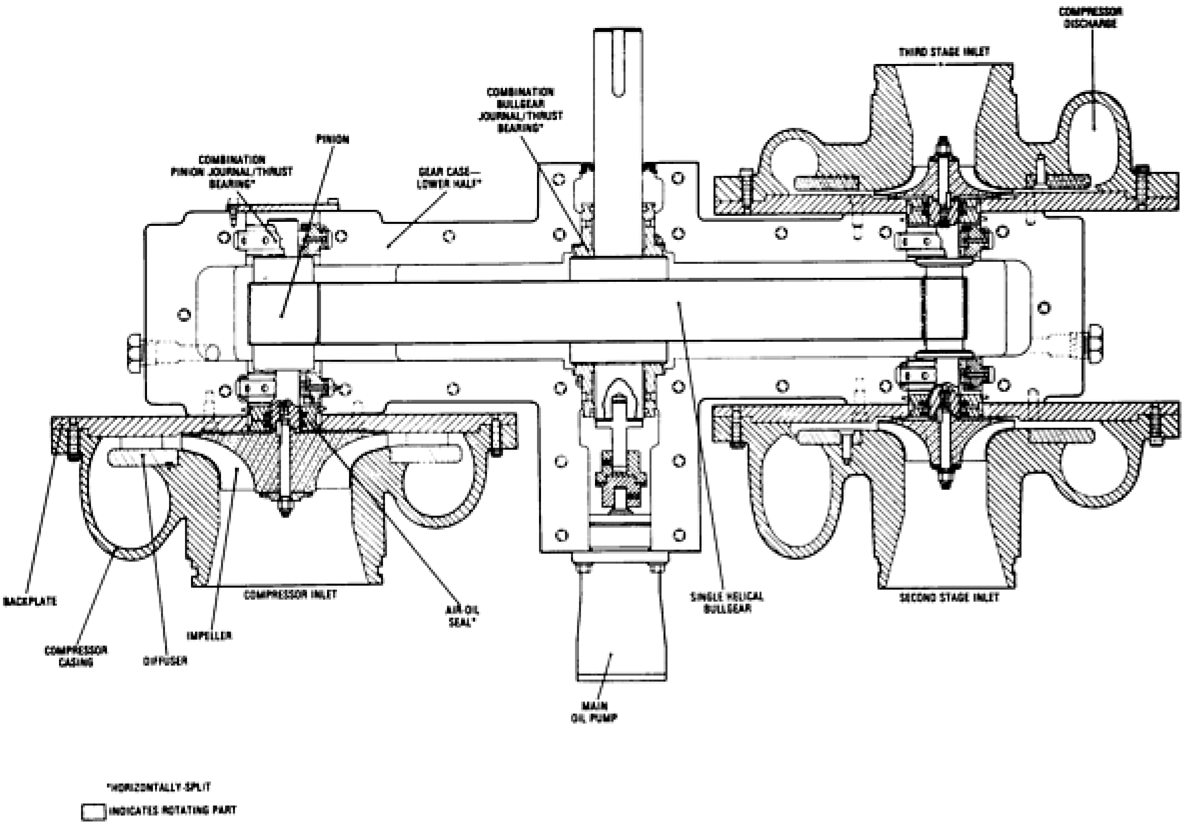


Fig. 13 Cross section of integrally geared centrifugal compressor. Courtesy of Elliott Company

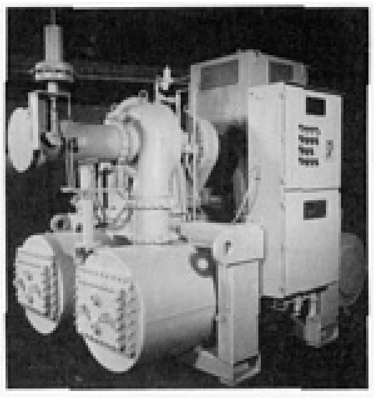


Fig. 14 Integrally geared centrifugal compressor. Courtesy of Elliott Company

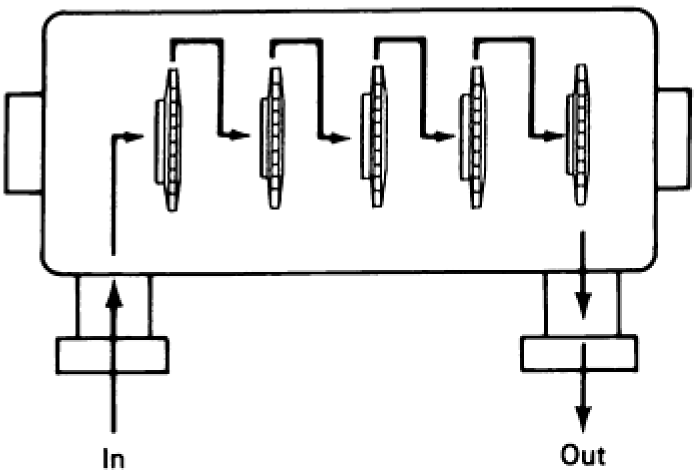


Fig. 15 Multistage centrifugal compressor with straight-through flow

Fig. 16 Horizontally split centrifugal compressor with straight-through gas path. Courtesy of Dresser-Rand

Another form of compounding is to use several impellers in series, take the gas out of the casing, cool it, and then return to the casing for more compression by a series of impellers (Fig. 17). This compounding can be repeated up to two times in one case.

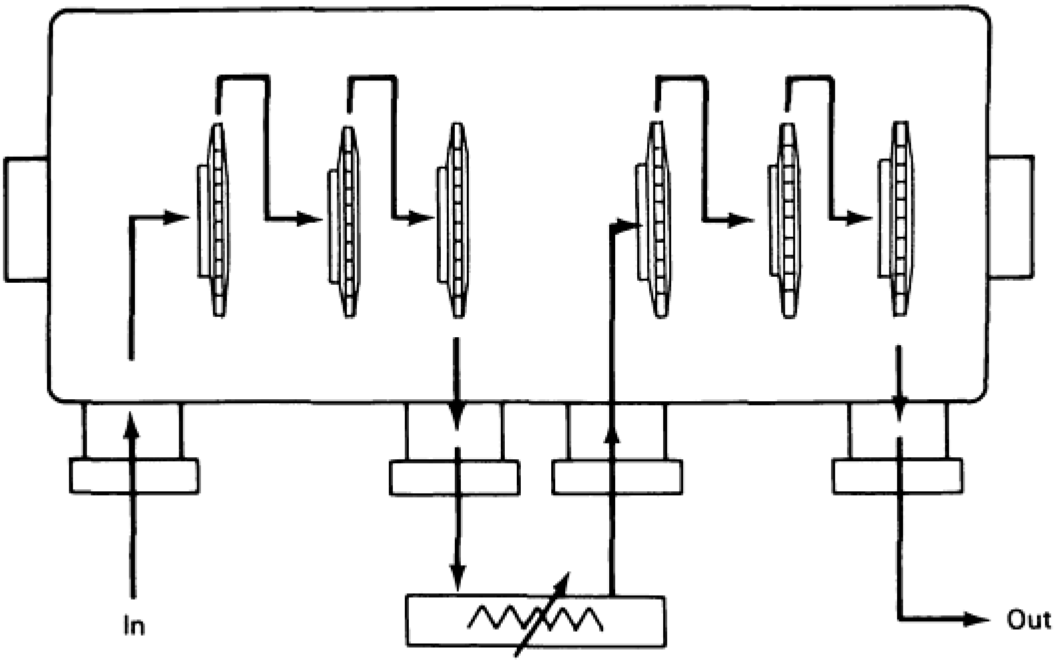


Fig. 17 In-out compound arrangement with intercooler

One common, somewhat complex configuration used in both process gas and refrigeration is the side-stream design (Fig. 18). The gas passes through one or more impellers, and is then joined by gas from an additional inlet at a compatible pressure level. This can be repeated several times. Gas can also be extracted from one of the side-stream nozzles.

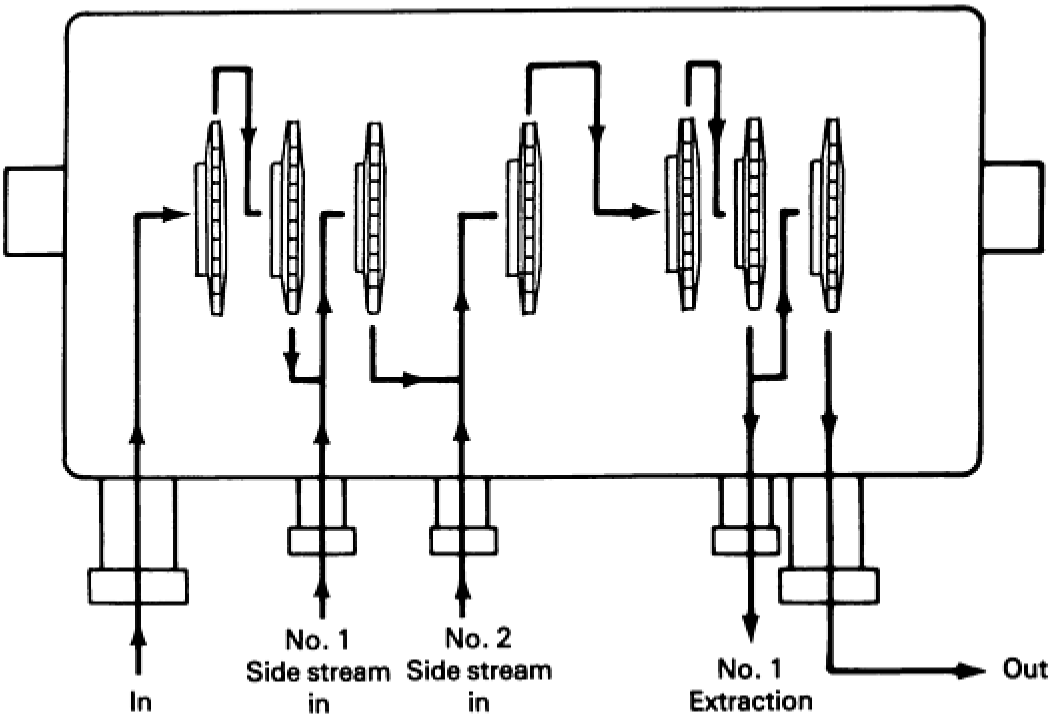


Fig. 18 Flow path through side-stream compressor

Bearings and Seals. The bearings used with a centrifugal compressor are of the oil-lubricated hydrodynamic type. Generally, the bearing designs are somewhat more sophisticated on the centrifugal compressor, because it operates at much higher speeds than the positive-displacement type. Rotor dynamics control is much more of a factor with the centrifugal compressor. Because the bearing is an integral part of the rotor support system, its dynamic characteristics must be known and controlled. The bearings can also contribute to rotor instability.

A bearing that has evolved for usage in the centrifugal compressor is the tilting pad bearing. It consists of a series of segments (pads) that are able to pivot in a manner that allows each pad to independently form an oil firm. Instabilities caused by oil whirl effects are thus eliminated. The thrust bearing is a tilting pad design that is symmetrical and carries an equal thrust load in either direction.

A new bearing design being used in commercial service is the magnetic bearing. Its advantage is that no lubrication is required, thereby eliminating a large external lubrication system. An external control cabinet for the bearing control is needed, but it is more modest in size. Because of the lowered friction that is due to magnetic levitation, the bearing losses are lower. With further development, the bearing will gain applications.

A disadvantage is that it cannot carry a unit load that is as great as a hydrodynamic bearing. This is normally not a problem for the compressor, but it is a problem for a gear or steam turbine. If a lubrication system is needed for the compressor driver, some of the advantage is lost.

Interstage seals are required on multistage (multiimpeller) configurations to minimize the leakage between one impeller and another, which has a direct impact on efficiency. It should be recalled that the labyrinth is a clearance seal that controls leakage by disrupting the flow with a series of teeth on the seal.

For air service, the labyrinth seal is also used as the shaft-end seal. A multiple labyrinth, using either injection or ejection, can be utilized with a limited number of gases.

More commonly, the gas compressors use either the carbon ring, liquid film, or the mechanical contact (buffered or dry) seal. These were described in terms of the helical lobe compressor and generally are of the same design.

Lubrication for the centrifugal compressor is provided by an external system. The lubrication system of a gas compressor in critical service is quite large. Although not as dramatic as the size difference described for the flooded helical lobe compressor, it sometimes is larger. One reason for its largeness is that the ability to maintain the lubrication system without shutting down the compressor must be allowed. Centrifugal compressors normally run for several years without shutting down. Three to five years is not unusual.

Except for the larger dry helical lobe compressors, this is not the case with positive-displacement compressors, whose lubrication systems must receive maintenance in that period. For this reason, almost all major components, pumps, coolers, and filters are normally duplicated. Space must be allocated on the skid to perform the maintenance.

Wear Considerations. Because of the nature of the centrifugal compressor, there are few, if any, true wearing parts. The bearings and seals operate with a clearance and, theoretically, do not rub. With oil-buffered seals, the improper cleanliness of the oil system will cause bearing and seal wear. With dry gas mechanical contact seals, the cleanliness of the buffer gas is a factor. Centrifugal compressors tend to foul in dirty service environments. The fouling can cause impeller vane wear. In some cases, liquid injection is used to combat the fouling. However, the cure can sometimes be worse than the malady, because the liquid can cause erosion and wear to internal parts.

This is also true in some applications where the gas is in vapor form near the saturation point. If too much liquid is allowed to form, erosion can occur. Surprisingly, the centrifugal compressor can absorb a misty liquid or a carefully controlled liquid injection very well. Labyrinth seals do tend to deteriorate with time. Although they are designed to not touch, start-up transients, process upsets, and products of fouling do cause wear.

Compressor Comparison

The reciprocating compressor and rotary compressor constitute positive-displacement types. The reciprocating compressor is the more complex machine, and is therefore more subject to problems from friction and wear. The rotary compressor is a much simpler compressor, with fewer wearing parts, but it does not have the pressure range capability of the reciprocating compressor. Applications are more limited, as well.

The centrifugal compressor is unique because it is a dynamic of compressor. Its application is in the higher-flow regime, and it can be used with a wide variety of pressures. The pressure ratio of a given compressor casing is not as high as that of the reciprocating compressor. The friction and wear characteristics are quite low, making the relative maintenance cost quite favorable.

Selected References

- R.N. Brown, *Compressors, Selection, and Sizing*, Gulf Publishing Co., 1986

Friction and Wear of Cutting Tools and Cutting Tool Materials

L. Alden Kendall, Department of Industrial Engineering, University of Minnesota, Duluth

Introduction

CUTTING TOOLS WEAR because normal loads on the wear surfaces are high and the cutting chips and workpiece that apply these loads are moving rapidly over the tool's wear surfaces. The cutting action and related friction at these constant surfaces increase the temperature of the tool material, which further accelerates the physical and chemical processes associated with tool wear. In order to remove the unwanted material as chips, these forces and motions are necessary; therefore, cutting tool wear is an economic penalty that must be accounted for in order to machine the part.

The magnitude of this economic penalty can be minimized if the cutting process is planned and controlled based on sound knowledge of the wear process and its dependency on the selection of cutting conditions. The cutting conditions normally controlled are the engagement of the cutting edge with the workpiece, the relative velocity of the cutting edge with the workpiece, and the feed velocity used to keep the tool engaged into uncut material. These cutting conditions are more frequently called depth of cut, cutting velocity, and feed rate. During the process planning stage, an assessment must be made concerning how difficult the material is to machine, the correct tool for the surface to be created must be selected, the appropriate tool material must be selected, and the type of cutting fluid needed must be determined. To make such choices, the wear environment in metal cutting must be understood.

The Wear Environment

Cutting tool wear occurs along the cutting edge and on adjacent surfaces. Figure 1 shows a view of the cutting process where the rake and clearance surfaces intersect to define the cutting edge. Figure 2 shows a similar view but adds the stress and strain states that are present in the material and chip. Figure 3 shows the force equilibrium and kinematic relationships that are commonly used to explain how these states of stress and strain are produced. The tool engages the material to depth t and the force vector **R** (**R** = **R**' = **R**") creates a component force, F_s, that shears the material along the shear plane area, A_s, to form a chip having thickness, t_c. The direction of the chip motion is controlled by the rake angle, α. The magnitude of the chip velocity, V_c, is a function of the cutting velocity, V, the rake angle, and the shear angle, ϕ. Relating Fig. 1 to Fig. 3, force F and the velocity V_c create the secondary shear zone, force F_s and velocity V_s acting along A_s create the primary shear zone, and force F_c and velocity V create the tertiary shear zone. These are idealized average forces and velocities. Cutting tools wear is localized on specific surfaces where stress, strain, velocity, and temperature are above critical levels. It is important to understand where these critical conditions exist and how they interact to cause tool wear.

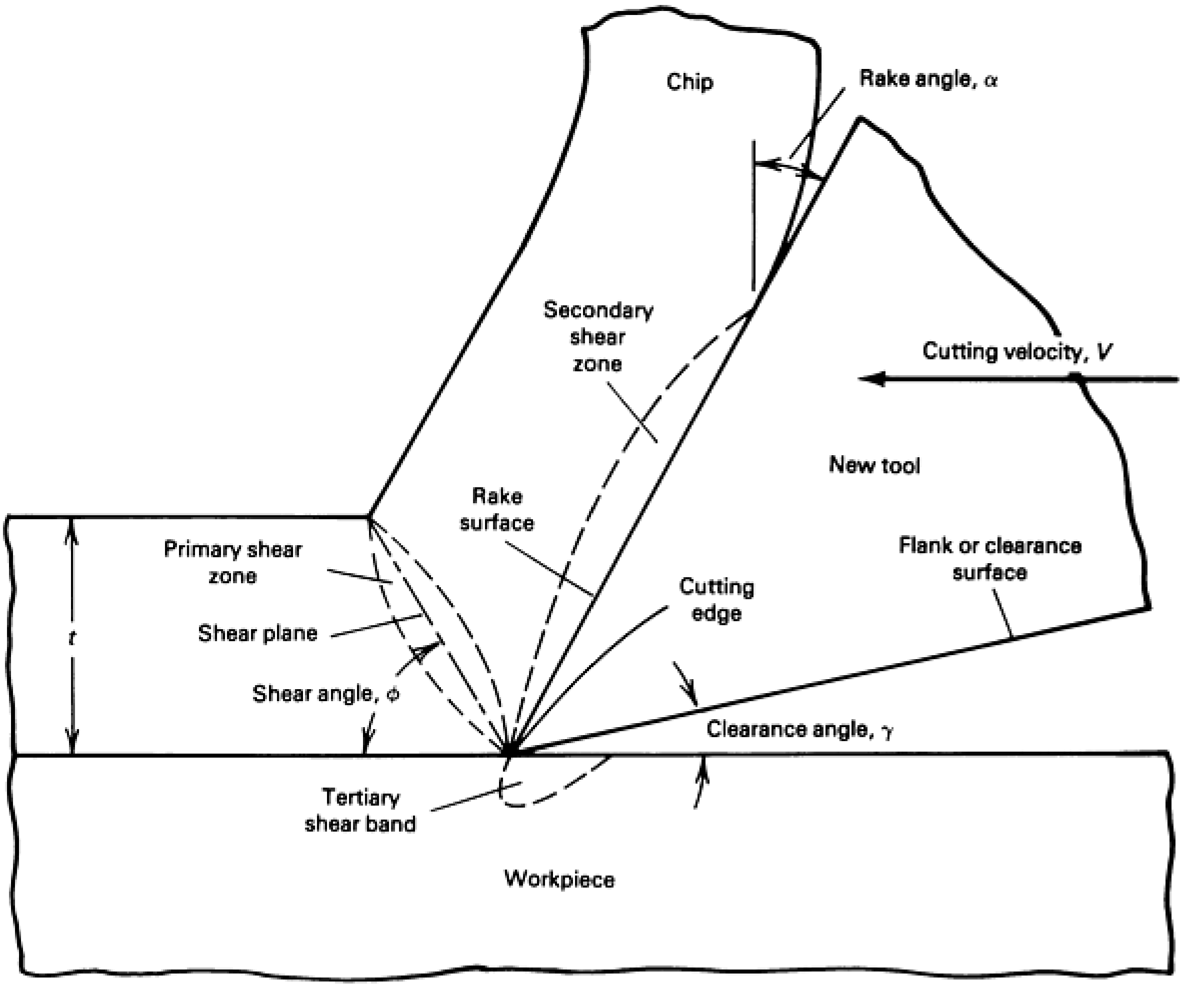


Fig. 1 Chip, workpiece, and tool relationship

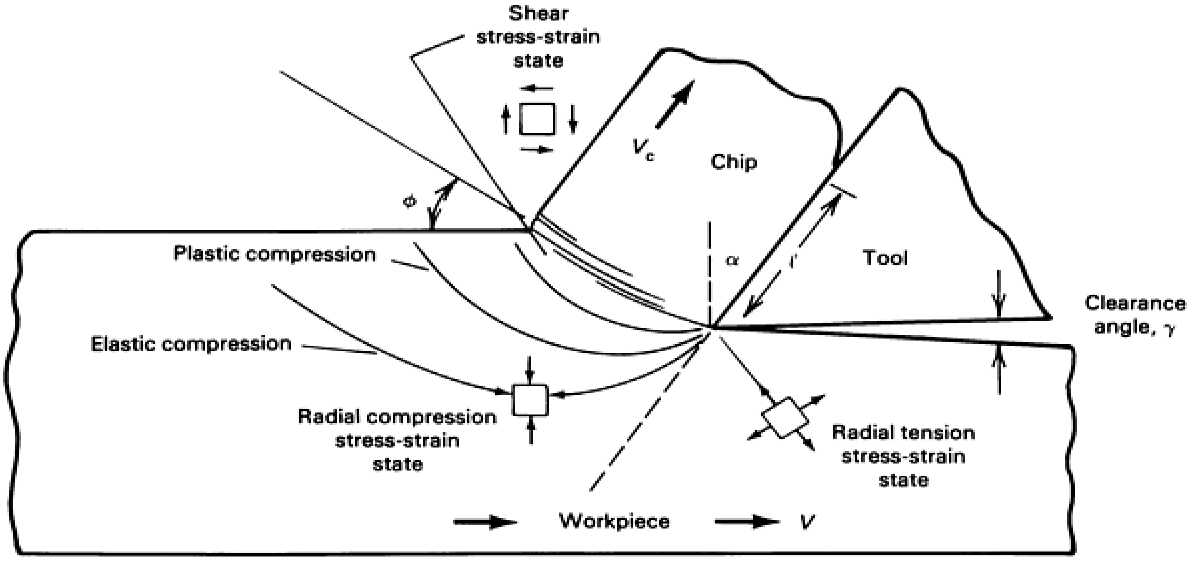


Fig. 2 Stress and strain states in cutting edge region. Source: Ref 1

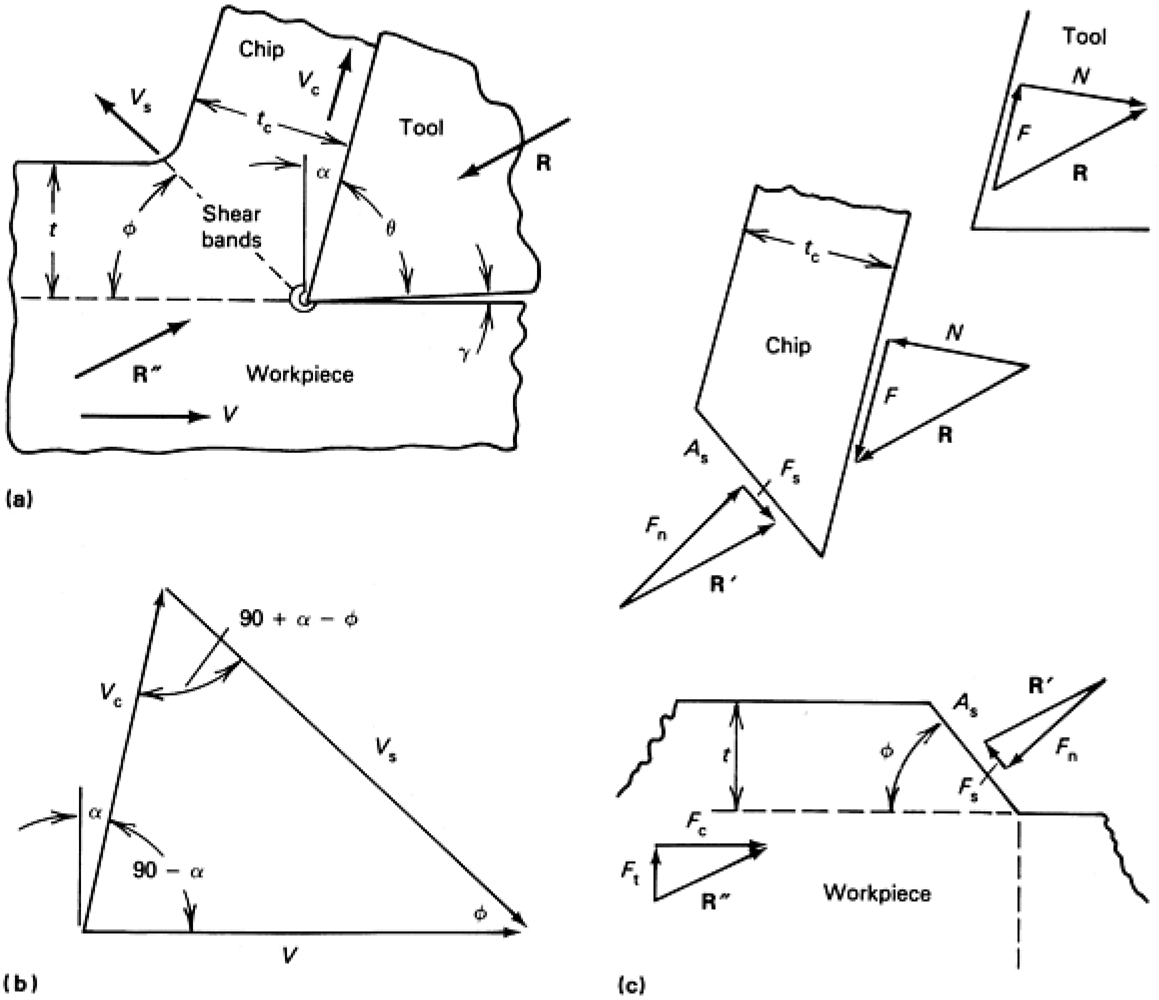


Fig. 3 Force equilibrium and kinematic relationships associated with orthogonal metal cutting. (a) Orthogonal model. *t,* uncut chip thickness; t_c, chip thickness; ϕ, shear angle; α, back rake angle; γ, clearance angle; θ, edge angle. (b) Velocity triangle, V_s, shear velocity; V_c, chip velocity; *V,* cutting velocity. (c) Force equilibrium freebody diagrams. *F,* friction force; *N,* normal to friction force; F_s, shear force; F_n, normal to shear force; F_c, cutting force; F_t, tangential force; R, force vector (R = R' = R"). Source: Ref 1

Wear Surfaces

Figure 4 shows how the sharp tool of Fig. 1 may wear due to the velocities and forces shown in Fig. 3. Along the rake surface, the chip motion, V_c, and high normal stress due to force *N* have produced a wear scar called crater wear. Along the clearance surface, the tool motion, *V,* and high normal stress due to force F_t have increased the area of contact between the tool and work, producing flank wear. Cutting edge wear has increased the radius and is caused by the combined effect of these component forces localized to the cutting edge region and represented by force vector **R** in Fig. 3. Figure 5 shows these characteristic wear surfaces on a turning tool insert, end mill, form tool, and drill. The cutting edge view shown in Fig. 1 and 4 is identified as section A-A in Fig. 5.

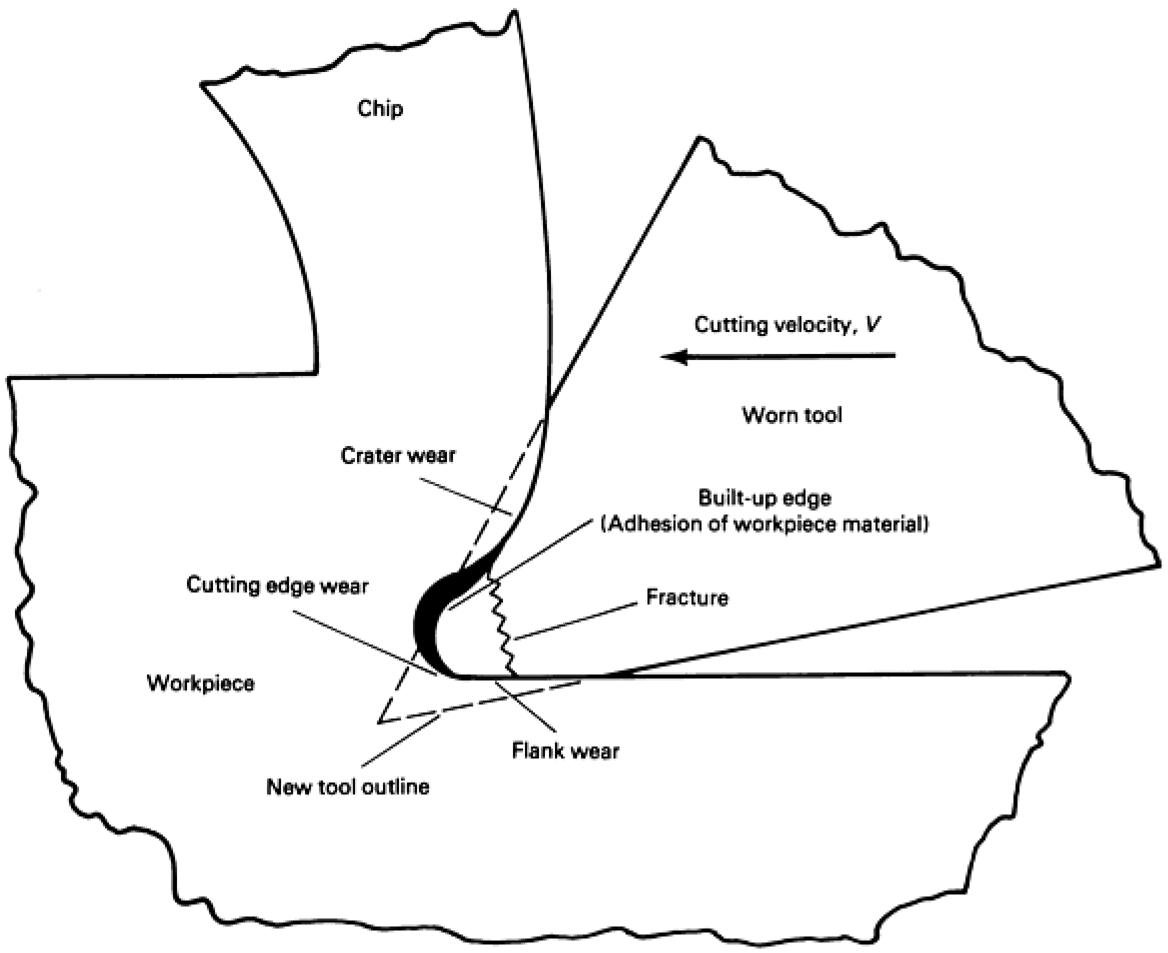


Fig. 4 Typical wear surfaces

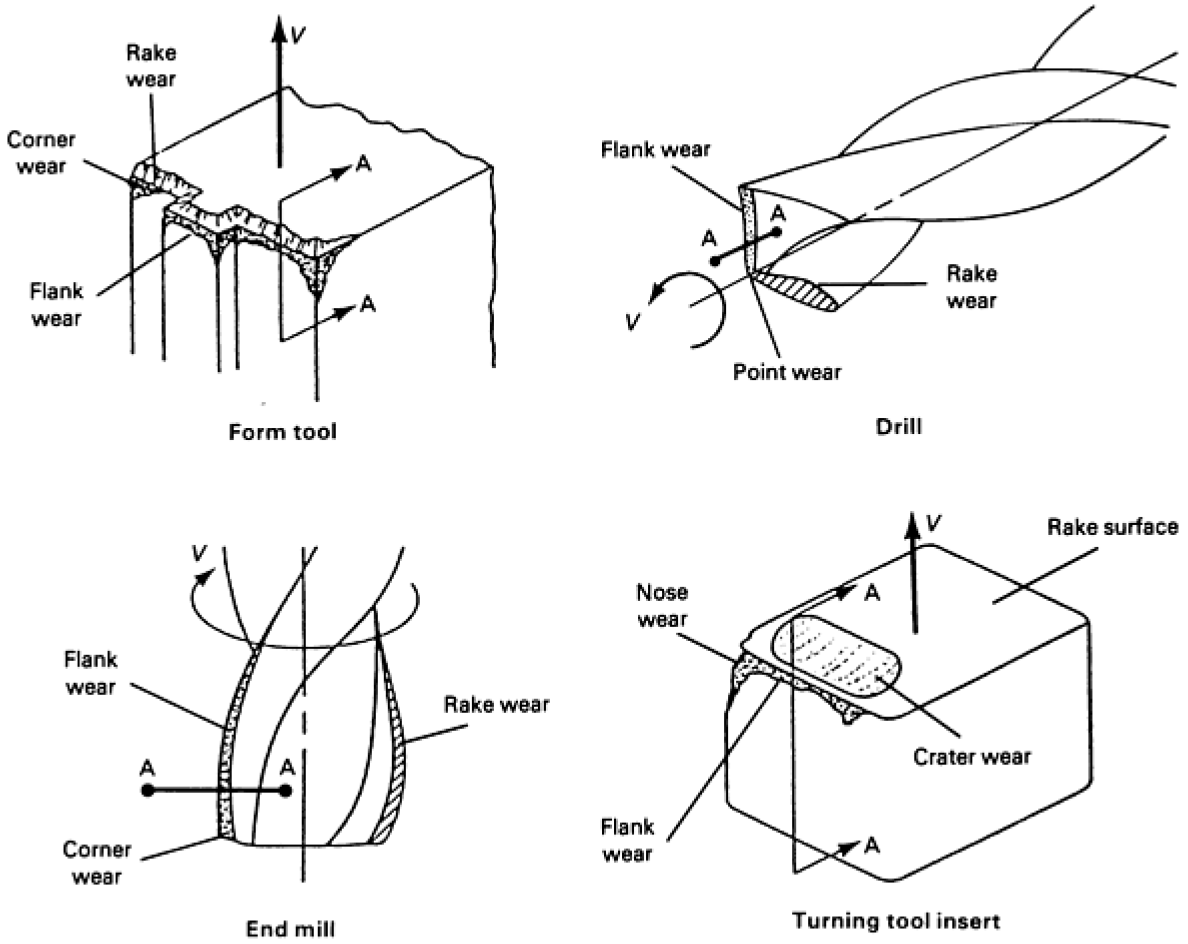


Fig. 5 Wear surfaces on common tools due to the tool motion, *V*

These figures show schematically how the wear process has changed the geometry of these different types of cutting tools. Flank wear decreases the diameter of the end mill and decreases the depth of cut for a lathe tool. Both of these changes in the geometry of the cutting tool could produce out-of-tolerance dimensions on machined parts. The edge wear and crater wear on the rake surface alter the state of stress and strain in the cutting region, thus changing cutting forces and the mechanics associated with the chip-making process. Severe geometric changes due to crater and flank wear can weaken the tool so that the edge may suddenly fracture as illustrated in Fig. 4.

It should be apparent that the location and size of these wear surfaces play an important role in determining the useful life of the cutting tool. Localized stresses on cutting tool surfaces are a major influence on the location and size of wear surfaces.

Stresses on Tool Wear Surfaces. Figure 6 shows the approximate distribution of normal and shear stresses on the tool's wear surfaces. The normal stresses, σ_n, are caused by the normal forces acting along the rake surface, the cutting edge surface, and the clearance surface. In addition to the normal stresses, Fig. 6 shows the shear stresses, τ, that act along the surface of the tool and that are associated with sticking and sliding shear processes. For the sticky zone, the normal force has a magnitude that results in a shear stress component which equals the shear yield strength, τ_y, of the strain-hardened work-piece material. Rather than sliding along the surface, the chip tends to adhere and periodically separate along shear fracture planes within the material, leaving adhered material on the tool or within the tool, which subsequently causes tool wear. The existence and size of this sticky zone are dependent upon the magnitude of the normal force and the frictional conditions along these surfaces. A sizable amount of this material may adhere for extended periods of time. Stable masses of such adhered material are known as built-up edge, and, as shown in Fig. 4, they alter the geometry of the cutting edge.

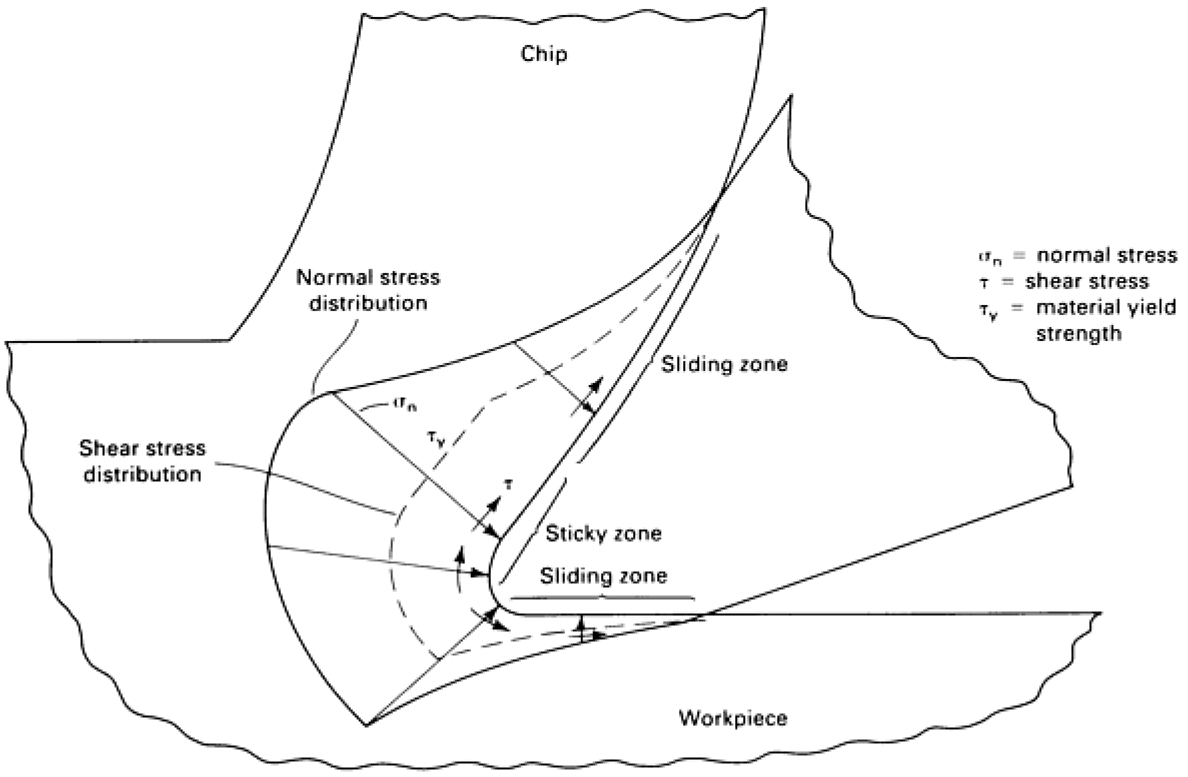


Fig. 6 Wear surface stresses

The sliding zones have friction forces and associated surface shear stresses that vary according to the normal force and coefficient of friction. Tool surface roughness and lubrication conditions affect the magnitude of these surface shear stresses. The state of stress created by the shear loads in the sticky and sliding zones along the rake surface produce the secondary shear zone shown in Fig. 1.

The primary shear zone in Fig. 1 extends from the cutting edge to the work surface; it is the zone where the work material is plastically deformed and sheared to form the chip. The complex state of stress along the cutting edge is caused by the strain associated with the chip separating from the work and moving along the rake surface of the tool and the strained material that remains a part of the work material. This conflict between chip motion strain and stationary workpiece strain produces a plowing action by the cutting edge.

Normal stresses can become very high and exceed the strength of the tool material, causing plastic deformation or fracture of the cutting edge. A sticky zone may not exist for certain cutting conditions; however, the plowing action along the cutting edge will always exist to some degree because it is impossible to create a cutting edge with no radius and a primary shear zone that is a perfect plane. The magnitude of the state of stress in the cutting region also varies with time, and the associated load variations may create a fatigue failure environment for the tool.

Motion along the Wear Surfaces. One way to increase machining productivity is to increase the volumetric chip removal rate, which is a function of the engagement area of the tool with the work material and the velocity at which the tool moves through uncut material. Therefore, one option the process planner has is to increase the cutting velocity. This productivity gain must be balanced against increased tool wear caused by higher cutting velocities.

Figure 3 shows that the chip velocity, V_c, and the shear velocity, V_s, are functionally related to the primary shear angle, ϕ, the rake angle, α, and the cutting velocity, V. The cutting velocity is the relative velocity between the clearance surface and the work, whereas the chip velocity is the relative velocity between the chip and the rake surface of the tool. The magnitude of these two velocities and the related shear stresses at these interfaces determines the amount of thermal energy released per unit of contact area. The velocity within the material adjacent to these two surfaces and the velocity, V_s, in the primary shear zone cause high local strain rates, and the volume of material strained by these stain rates releases additional thermal energy. These thermal energy releases cause the temperature of the workpiece, cutting tool, and chip to

increase. The productivity gain from increasing the cutting velocity proportionately increases wear surface velocities, strain rates, and the release of thermal energy, thus increasing the wear environment temperature.

Temperatures in the Wear Zone

The difference between the thermal-energy release rate and the thermal-energy dissipation rate determines the temperature of the material and tool in these wear zones. Thermal-energy dissipation is a function of the thermal conductivity properties of the tool material and workpiece material. The chip carries some of this thermal energy away from the tool and work material. Additionally, the workpiece size and specific heat determine the workpiece heat capacity, and, to a lesser extent, the surface area plays a role in convective and radiative heat transfer to the surrounding air. In a cutting fluid is used, the rate at which it removes thermal energy from the hot tool and workpiece plays an important role.

Tool materials are developed and selected based on their ability to maintain their hardness, toughness, and chemical stability at high temperatures. Even with the best of tools, such material properties eventually adversely change with increasing temperatures. There have been many studies of temperatures in the cutting tool. Most commonly, average temperature conditions are approximated using the following type of relationship (Ref 2):

$$T = u\left(\frac{Vt}{k\rho c}\right)$$

where *T* is the mean temperature of the tool-chip interface, *u* is the specific cutting energy (which is the energy used per unit volume of material removed), *V* is the cutting speed, *t* is the undeformed chip thickness (see Fig. 1), and *k*, ρ, and *c* are the conductivity, density, and specific heat, respectively, of the workpiece material.

Figure 7 shows a typical temperature distribution. The high-temperature gradient can cause portions of the tool to be at dangerously high-temperature levels, well above the mean predicted temperature. The mean temperature relationship given above does identify the important variables and how they affect temperature. High-strength work materials have high values of specific cutting energy. Some of these high-strength materials, such as titanium, are also poor conductors and are of low density. This combination of conditions results in higher tool temperatures, and the cutting tool material used must maintain its hardness and strength at these temperature levels if it is to have a useful service life. Combining these temperature effects with the state of stress, strain rates, and material motion makes the wear zones of the tool a complex battleground in which these conditions interact to trigger the mechanisms that cause wear.

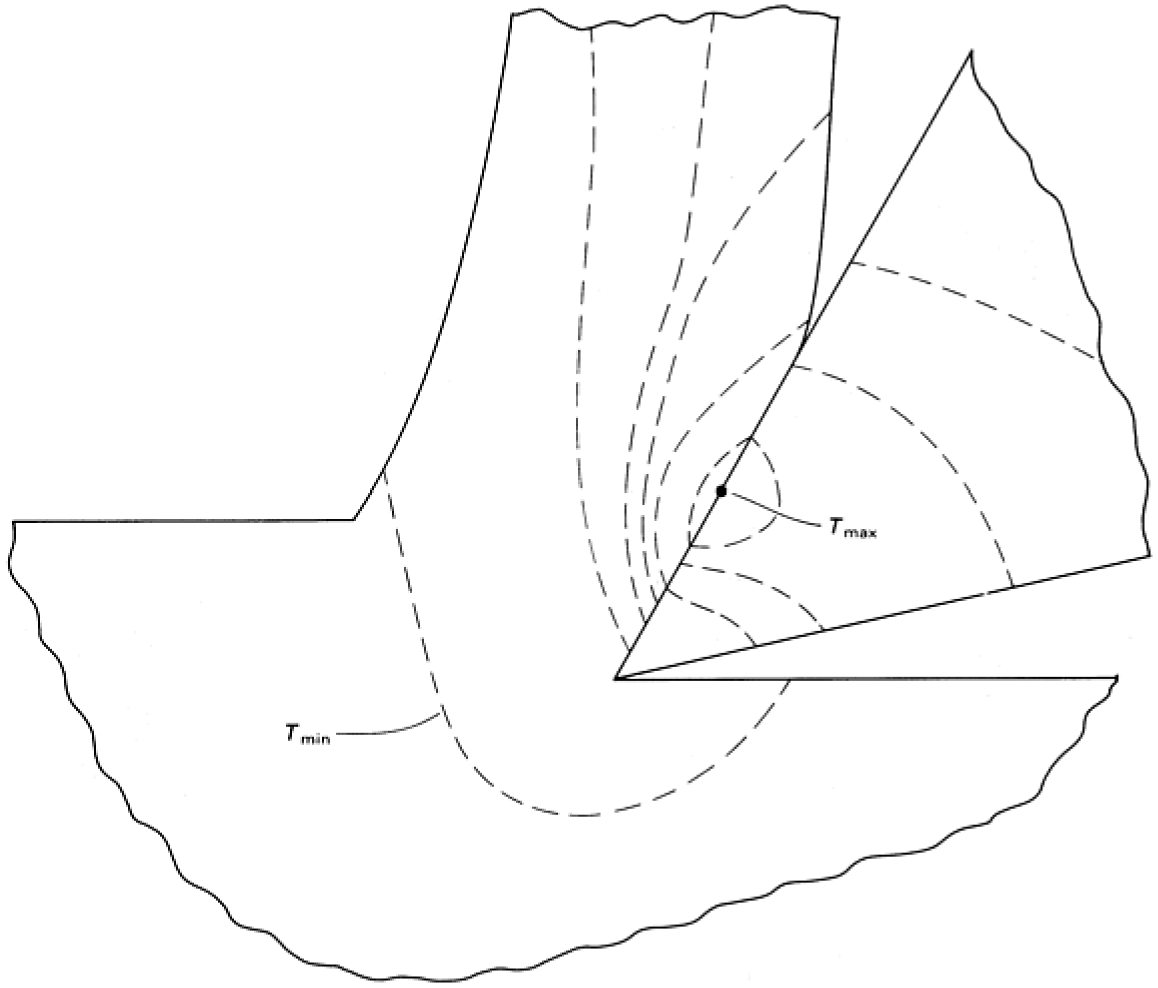


Fig. 7 Generalized temperature distributions in the cutting region

Wear Mechanisms

One important area of concern in wear studies has been the identification of wear mechanisms. What has emerged is that a particular wear mechanism is dependent upon the contact stress, relative velocities at the wear interface, temperature, and physical and chemical properties of the contacting materials. For a particular set of contacting materials, wear mechanism maps have been used to identify the ranges of normal pressure and velocity that result in a particular wear mechanism (Ref 3). Figure 8 shows the general structure of such a map when the area of the contacting wear surface remains constant. Temperature is not a mapped variable, however, it is a variable that depends upon the pressure (normal stress), velocity, and size of the wear surface as discussed in the previous section of this article. The four broad classes of mechanisms shown in Fig. 8 are:

- Seizure
- Melt wear
- Oxidation/diffusion-dominated wear
- Plasticity-dominated wear

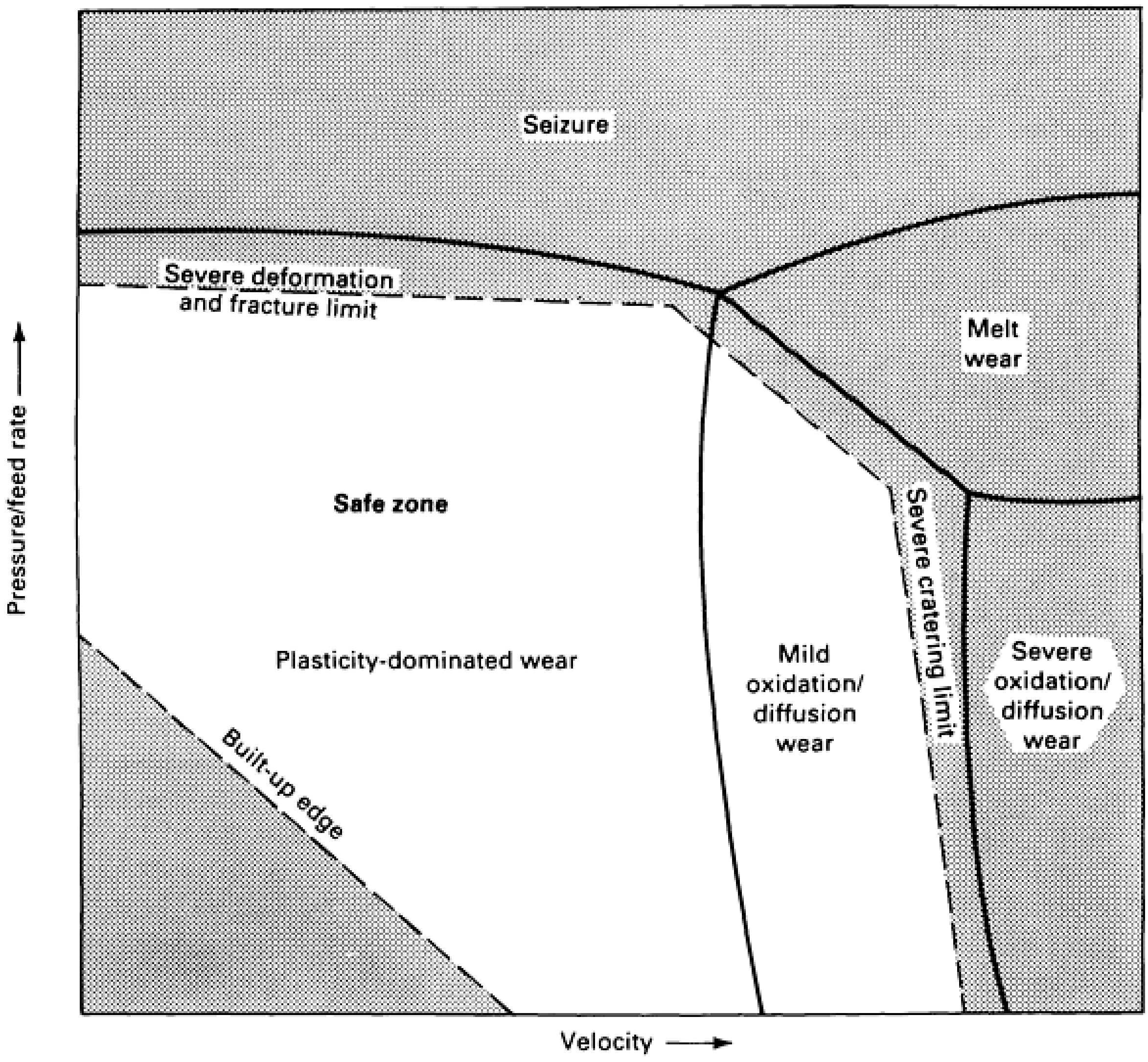


Fig. 8 Wear mechanism maps and safe operating regions for cutting tools. Source: Ref 3

Cutting tool wear research has established similar maps for feed rate, which is related to pressure for a given engagement condition, and velocity (Ref 4). The dashed lines in Fig. 8 show some possible boundaries that could be used to define a safe operating zone. These boundaries provide a framework for discussing the progressive wear and failure conditions for cutting tools.

Initial Wear Mechanisms. The two materials in contact have surface roughness irregularities in the form of protrusions or asperities. At the interface, asperities from the two materials touch, defining tiny contact areas. The total area from these contact points is a fraction of the projected area of the contact surface. The stresses and heat are intensified in the asperities, and partial removal may occur due to seizure accompanied by fracture of the asperity or melting in the asperity. As these asperities are removed, the initial surface roughness is altered and the contact area increases. If the force conditions remain unchanged, pressure decreases and the active wear mechanisms change to plasticity and/or mild oxidation/diffusion-dominated wear. This initial wear period will create small, visible wear surfaces.

Steady-State Wear Mechanisms. A velocity and normal stress condition that would continue to cause seizure and melt should be avoided because it would soon cause complete failure of the cutting tool. Assuming that such conditions do not exist, the wear surfaces will get progressively larger. If plasticity dominated, small particles of material are mechanically deformed and fractured away from the wear surface. This is generally called abrasion and can occur on any of the wear surfaces; it is the most common wear process along the clearance or flank surfaces of most tools. As discussed

earlier, normal stress and temperature vary over the wear surfaces so that a plasticity mechanism that dominates in one wear zone may not dominate in another.

Figure 7 indicates that the maximum tool surface temperature occurs on the rake surface a small distance from the cutting edge when cutting most materials. This is where the crater wear condition occurs; diffusion wear is often the dominant mechanism (Ref 5). The high temperature and pressure cause atoms to move between the contacting materials, and this diffusion process locally aids in the removal of material to form the crater. In very hard cutting tool materials, such as ceramics, where high cutting velocities are commonly used, these oxidation and diffusion mechanisms may be responsible for a majority of the wear.

The dashed line in the lower left-hand corner of Fig. 8 that identifies a built-up edge (BUE) condition affects the process in two ways. Near the cutting edge, the higher pressures could cause particles of work material to adhere to the cutting tool in the sticky zone. If the shear forces due to chip movement are high enough, the bond will be temporary and the adhered material will fracture away from the tool surface. When it fractures, small particles of tool material may be removed with the previously adhered material. This then is a wear process and would be associated with conditions in the safe zone just outside the BUE line. The second effect on the process occurs when the BUE is not fractured away by the chip motion and remains to alter the geometry of the cutting edge. The presence of the BUE changes the shear angle, causing instabilities in the chip-forming process plus damage to the machined surface. The lubricating characteristics of cutting fluids are helpful in eliminating this BUE condition. For rough cuts, the presence of a stable BUE can be beneficial in that it provides an intermediate protective layer between the tool surface and the work material.

Wear can also occur as chipping along the cutting edge. Such chipping more commonly occurs when the cutting edge intermittently removes chips. This results in cyclic impact and thermal loading to the cutting edge. These two cyclic loading states can initiate small cracks and then cause these cracks or other residual cracks to propagate, leading to tool chipping.

These abrasion, oxidation, diffusion, and chipping wear mechanisms that occur at operating conditions within the safe zone shown in Fig. 8 cause the initial wear surfaces to gradually enlarge over time. This surface life period is often referred to as the steady-state wear period.

Tertiary Wear Mechanisms. The steady-state period of wear will eventually enlarge the wear surfaces to a critical size that will trigger accelerated wear. The pressures and velocities on these enlarged surfaces begin increasing the temperature such that the rapid oxidation/diffusion and local seizure or melting conditions cause rapid destruction of the tool. In tools that have a hard, wear-resistant coating such as titanium nitride, wear through this coating or separation of small volumes of the coating from the base material will expose the less-resistant core material to these same rapid wear conditions. A tool change must be made before this point is reached. Figure 9 shows these three zones in terms of the amount of wear over time.

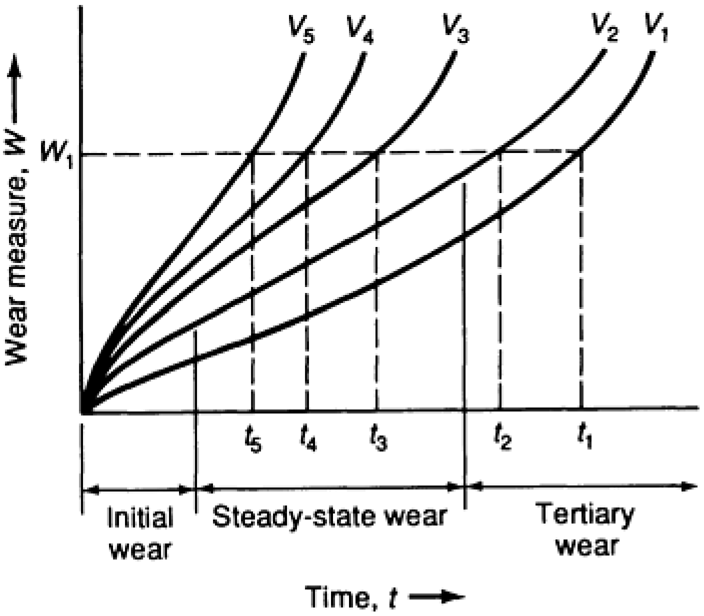


Fig. 9 Tool wear curves for different cutting velocities

To further complicate the understanding of cutting tool wear mechanisms, there is a growing trend toward high-speed machining. This means using cutting velocities 5 to 10 times the normal speed. New tool materials and machine tool designs have made this possible; however, these changes after the wear environment. Proportionally more thermal energy is removed with the chip due to the decrease in contact time between the tool and the chip; however, the higher velocities will increase absolute temperatures on the tool wear surfaces. Under such conditions, abrasive wear may become less important and the diffusion and oxidation processes could dominate in creating and enlarging the wear surfaces.

Machine, Cutting Tool, and Tool Wear Interactions

The cutting tool is only one element of the machining system. This is a complex system consisting of the workpiece, the workpiece holding fixture, the tool holder, the machine tool, and the cutting fluid system. Also, more complex cutting tool geometries are now available that control chip motion or chip breakage such that the process is more productive and the wear rate of tools reduced. When configuring a machining system, the selection process starts by obtaining knowledge about the type and quality of the surfaces to be machined. These surfaces, as well as the material to be machined, are specified as inputs to this configuration process. The geometry of these surfaces will determine the types of cutters that are needed.

The selection of the machine tool depends upon the sequence of surface types that will be used, the quality of each of these surfaces, and machine capacity needed for each of these surfaces. Machine capacity refers to the largest part dimensions, the cutting parameters, maximum power requirements, and the force loads generated by the cutting processes. To determine this machine capacity, tool materials and cutting fluid conditions must be established.

To simplify the discussion, consider a situation where a single type of surface is being machined to a given quality level (that is, tolerance and surface finish). The selection of the machine tool, the type of cutting tool, the type of cutting tool material, the cutting fluid conditions to be used, the feed rate, the cutting velocity, and the depth of cut defines one operating state for the machine system. This state will establish the wear environment at the cutting edges of the tool. How rapidly this wear environment degrades the tool will determine how frequently the tool will have to be changed. A tool change interrupts the machining process, which subsequently decreases the production rate and adds a cost that is associated with the interruption, the tool change expense, and the replacement tool cost. Trying to reduce this cost by lengthening the interval between tool changes will eventually cause the cost to rise again. This rise occurs when the wear process has progressed to induce tertiary wear mechanisms and accelerated wear of the cutting edge. This accelerated wear damages the surface, necessitating repair expense or the costs associated with a scrapped part. Even if the tool-changing interval keeps the wear process within the steady-state wear zone, there is a finite probability that the tool may

fail unexpectedly before a scheduled replacement and cause damage that requires repair or scrapping of the part. How probable such an event is depends upon workpiece material characteristics, the nature of the cutting loads on the tool, and the quality of the cutting tool. Note that without changing the state of the machine system, a fairly complex sequence of decisions has to be made to establish a tool-change strategy. For each new state of the machining system that is examined, a new tool change strategy will result. This search process for the best operating state is not an objective of this article; Ref 6, 7, 8, 9, and 10 describe the methodologies associated with establishing an interactive machining system. The remaining sections of this article focus on the following important factors involved in mitigating cutting tool wear:

- The basic guidelines for selecting tool materials
- Testing of cutting tools to provide information to guide the selection process
- Selecting cutting fluids

Cutting Tool Material Selection Guidelines

Tool materials need to be wear resistant, tough, and have a characteristic referred to as hot hardness. Wear resistance is the ability of the material to have a useful life when subjected to the types of wear mechanisms discussed earlier. Because the type of active wear mechanism varies with the wear environment, tool manufacturers have a broad spectrum of materials available that are wear resistant when used for different cutting conditions (Ref 11, 12, 13). The common tool material categories are high-speed steels (HSS), cast cobalt alloys, micrograin (powder metallurgy) high-speed steels, and various grades of carbides, ceramics, polycrystalline diamond (PCD), or cubic boron nitride (CBN). Some of these materials are further coated with harder material such as TiN, TiC, or Al_2O_3.

Toughness is needed so that the cutting edge will not chip or fracture, particularly when subjected to impact loads. Tough materials can absorb energy and resist plastic deformation without fracturing. As a general rule, increasing hardness decreases toughness. Figure 10 compares these classes of tool materials for toughness and wear resistance.

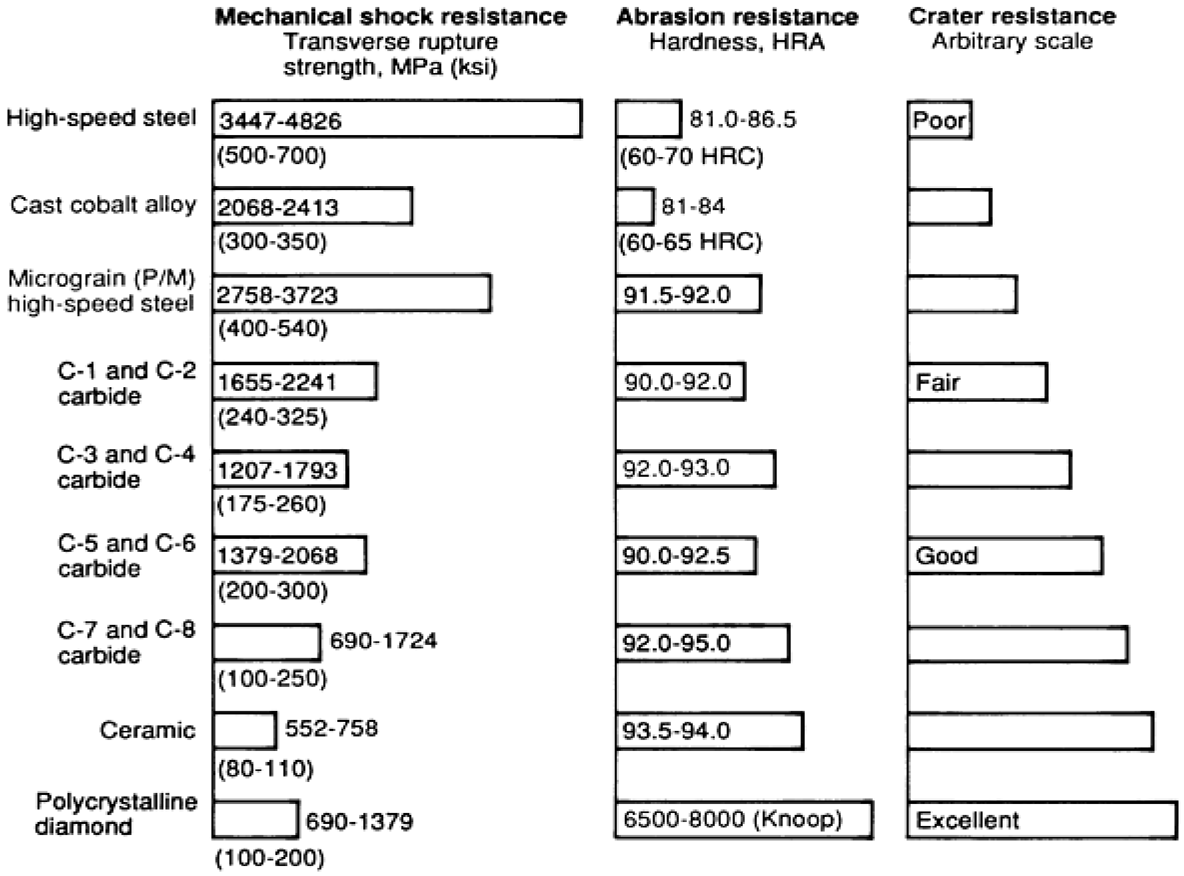


Fig. 10 Toughness and wear resistance of various classes of cutting tool materials. Source: Ref 13

Hot hardness is the ability of the material to maintain hardness at elevated temperatures. Recovery hardness is the ability of the material to regain hardness at room temperature after being subjected to an elevated temperature. Figure 11 compares hot and recovery hardness for various classes of tool materials.

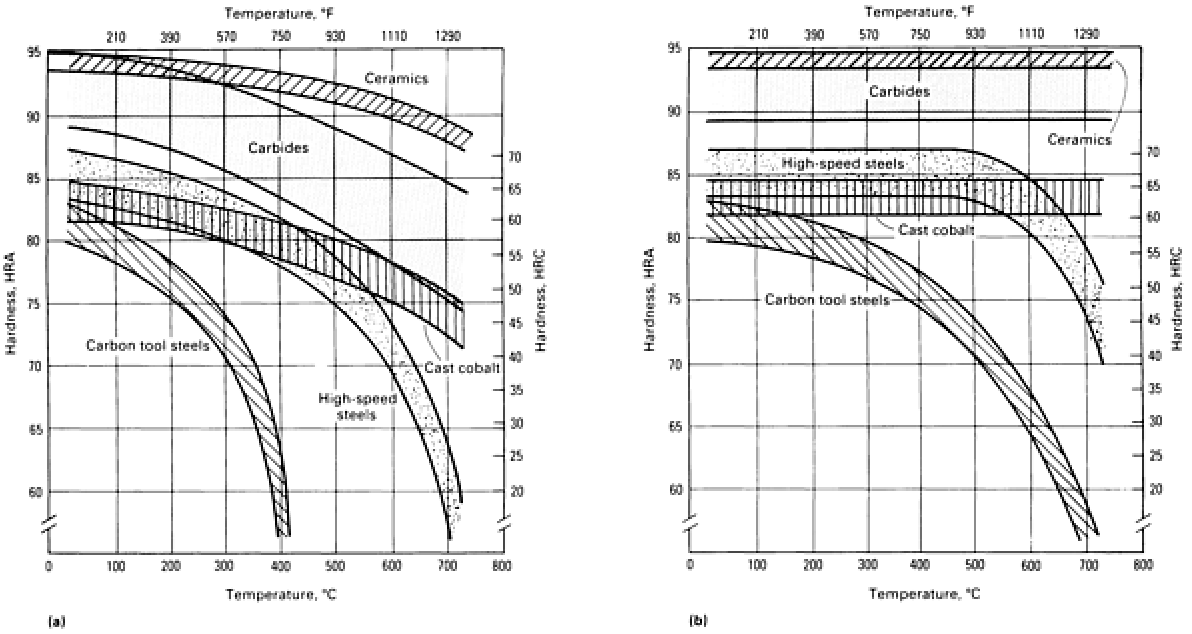


Fig. 11 Hot hardness (a) and recovery hardness (b) for cutting tool materials. Source: Ref 13

To select the correct cutting tool material, the process planner may need to depend upon the expertise of the tool supplier or the work material supplier, on experience, or on the results of cutting tests. The recommendations of tool and material vendors tempered by information gained through experience is the least costly way to make the selection. If the quantity of certain types of cutting tools used is large, the expense of a cutting tool evaluation program can be justified.

As already emphasized, the selection of the tool material is dependent on the type and quality of the surface being machined. A general rule connected with this selection is that a finish cut will have a small depth of cut but a large cutting velocity. The small depth of cut reduces the tool loads, whereas the high velocity may create a harsh wear environment. In this instance, toughness requirements are reduced and wear resistance plus hot hardness requirements increase. If the surface is a rough machined surface, large depths of cut and lower cutting velocities are used. This combination of conditions requires a tool material that is tough but that may not need the high level of wear resistance required for higher-velocity cutting conditions. Cutting tool manufacturers have developed elaborate systems to more specifically classify their cutting tool materials; however, their systems are often based on some logical trade-off between wear resistance and toughness for different levels of tool loading caused by different cutting edge engagements, feed rates, and cutting speeds. Figure 12 relates these trade-offs to tool failure mechanisms. Knowing that this sort of trade-off is being made makes it possible to better evaluate vendor recommendations.

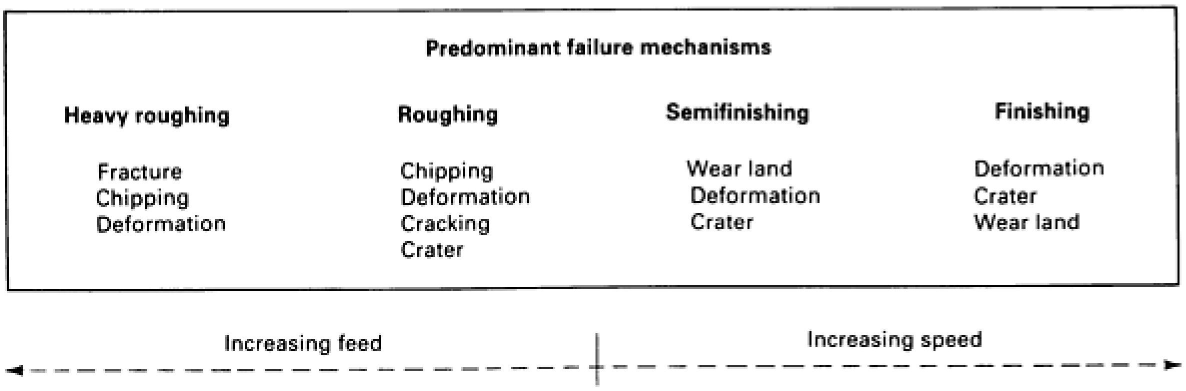


Fig. 12 Failure mechanisms as a function of engagement, feed rate, and cutting speed

When selecting a cutting tool material, a recommended cutting speed range is established. The type of wear mechanism that will likely exist is determined by the cutting speed and the type of material being machined using a particular cutting tool material. This relationship between work material and cutting tool material is discussed later; however, Fig. 13 summarizes some of these relationships. Some general guidelines for selection of cutting tool materials based on the type of machining operation are provided in Table 1. Again, the dependence upon work material is included.

Table 1 Cutting tool material selection based on machining operation and work material

Machining operation	Work material	Applicable tool material						
		High-speed steel	**Carbide**	**Cast alloy**	**Coated carbide**	**Ceramic/ cermet**	**Cubic boron nitride**	**Diamond compacts**
Turning	Cast irons, carbon steels, and alloy steels	X	X	X	X	X	. . .	. . .
	Alloy cast irons		X	X	X	X	X	. . .
	Aluminum, brass	X	X	X	X	. . .	. . .	X
	Nickel-base alloys, titanium	X	X	X	X	X	. . .	. . .
	High-silicon aluminum	. . .	. . .	. . .	X	. . .	. . .	X
	Plastics, composites	. . .	X	X	. . .	. . .	X	. . .
Milling	Cast irons, carbon steels, and alloy steels	X	X	X	X	X	. . .	. . .
	Alloy cast irons		X	X	X	X	X	. . .
	Aluminum, brass	X	X	X	. . .	. . .	. . .	X
	Nickel-base alloys, titanium	. . .	X	X	. . .	. . .	. . .	X
	High-silicon aluminum	. . .	. . .	. . .	. . .	. . .	. . .	X
	Plastics, composites	. . .	X	. . .	. . .	. . .	. . .	. . .
Drilling, reaming, tapping	Cast irons, carbon steels, and alloy steels	X	X	. . .	X	. . .	. . .	. . .
	Alloy cast irons	X	X	. . .	X	. . .	. . .	. . .
	Aluminum, brass	X	X	. . .	. . .	. . .	. . .	. . .
	Nickel-base alloys, titanium	X	X	. . .	. . .	. . .	. . .	. . .
	High-silicon aluminum	. . .	X	. . .	. . .	. . .	. . .	. . .
	Plastics, composites	. . .	X	. . .	. . .	. . .	. . .	. . .
Gear cutting, hobbing, broaching	Cast irons, carbon steels, and alloy steels	X	X	X	. . .	. . .	. . .	. . .
	Alloy cast irons	X	X	. . .	. . .	. . .	. . .	. . .
	Aluminum, brass	X	. . .	. . .	. . .	. . .	. . .	. . .
	Nickel-base alloys, titanium	X	X	X	. . .	. . .	. . .	. . .
	High-silicon aluminum	X	. . .	. . .	. . .	. . .	. . .	. . .
	Plastics, composites	X	. . .	. . .	. . .	. . .	. . .	. . .

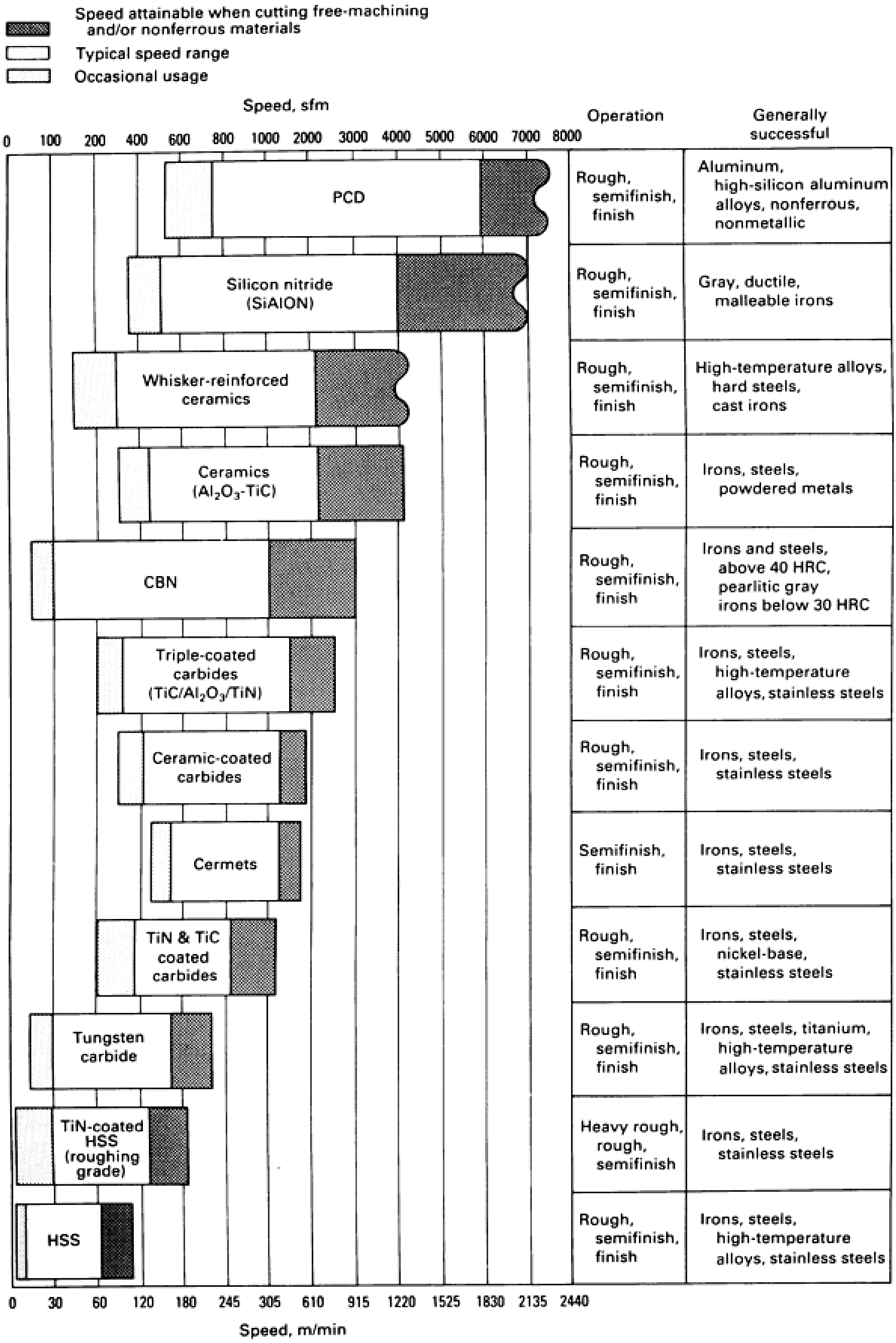


Fig. 13 Relationship between classes of cutting tool materials, cutting speed, and work material. Source: Ref 13

The type of cutting tool and machine tool influence what type of cutting tool material can be used. The versatility needed from the tool and the complexity of the geometry of the tool are such determining factors. An older machine tool with lower speed and power may not have the capability and rigidity to effectively use tools with a higher cutting speed capability. In order to provide some guidance, the following sections describe typical applications for the various types of cutting tool materials. More detailed information on the processing, properties, and applications of these materials can be found in the Selected References supplied at the end of this article.

High-Speed Steel (HSS). In the past, a majority of the machining operations used tools made from this material. The need for higher production rates, however, is resulting in more tool products using more wear-resistant materials. Reamers, taps, thread chasers, gear cutters, saws, and complex form tools use HSS exclusively. Milling cutters and drills are frequently made from this material. High-speed steel is also still used for a wide variety of turning tools because it is inexpensive and performs well when machining at moderate cutting velocities on many of the more machinable materials. It is the easiest material to grind to the desired geometry and has good impact strength compared to the other harder cutting tool material alternatives. This makes HSS very good for tools used for interrupted-cut, low cutting velocity machining operations on difficult-to-machine materials such as heat-treated steel, titanium alloys, and high-temperature materials.

High-speed steel comes in a range of grades that provide various combinations of wear resistance, hot hardness, and toughness. High-speed steel tools can also be coated with hard carbides or nitrides to provide tools with higher wear resistance that still maintain the toughness of the HSS core material.

Cast Cobalt Alloys. Tools cast from cobalt-chromium-tungsten alloys have high hot hardness, are reasonably tough, and have good wear resistance. This material is often a good choice for machining conditions between the conditions generally used for HSS and those used for tungsten carbide. It is also a good material for tooling used in a multiple-tool setup in which the variety of cutting conditions used for the machining operations require that the tool perform well at low and high cutting speeds as well as in a variety of tool-loading states. Typical applications for cast cobalt alloys include solid tools cast to shape, cutting blanks that are brazed to the tool, and cast inserts that are mechanically fastened to a tool holder. Turning tools, drills, mills, and cut-off tools can all use these blanks and inserts for cutting edges.

Micrograin high-speed steels are made using powder metallurgy (P/M) processing techniques. The atomized powder is extremely fine and of uniform composition; no binders are used. Either hot isostatic pressing (HIPing) or vacuum sintering can be used to produce a dense and uniform product. The material can be further heat treated in the same manner as wrought HSS. The microstructural uniformity (carbide distribution) of P/M tool steels and the absence of metallurgical defects make the material tough, and the HIPed or sintered shapes retain their dimensions when heat treated to the desired hardness level. Increasing alloy content to improve wear resistance is possible without severely affecting toughness. Almost all types of tools are made from this material; thus, P/M tool steels should be seriously considered if the machining benefits justify the increased cost over a competing material.

Cemented Carbides. A variety of carbide materials are available in fine powder form and can be produced using P/M processing into a coherent, hard material using binders such as cobalt. The variations and grades that are now commercially available are too numerous to discuss in detail (Ref 13, 14). The development of coated carbides has led to an increased range of applications. The type of powder, binder, and powder particle size determine the final properties. Cemented carbide cutting tool suppliers provide a broad range of grades, from those suitable for finished surfaces using high cutting speeds and small depths of cut to tougher grades more suitable for the lower cutting speeds and heavy tool loads found in rough cuts.

Carbide inserts come in a variety of sizes and are mechanically clamped to the tool at a position and orientation that is needed for the tool to effectively remove the material and create the required surface geometry. Manufacturers have found ways to clamp these inserts on many types of turning tools, mills, and drills. Such inserts usually have three to four cutting edges; when one edge is worn, the insert can be unclamped, indexed to a sharp cutting edge, and reclamped. After all edges are worn, the insert is not resharpened but instead thrown away. This indexability feature has made cemented carbide inserts very popular in production environments because they reduce the cost of tool replacement while achieving high material removal rates.

The hardness of these insert materials makes them resistant to abrasive wear, but vulnerable to chipping and other forms of fracture wear. Cyclic thermal loading can induce fatigue cracks. The clamped carbide insert conducts heat poorly, and generous use of coolants is needed to prevent thermal load damage. They have high hot hardness and can cut effectively at high temperatures. For certain work materials, temperature conditions, and cutting surface pressures, diffusion and

oxidation wear processes very rapidly degrade the cutting edge. In such applications, ceramic inserts may be more effective.

Ceramic inserts can be made from aluminabase or silicon nitride-base materials, or they may have 15 to 39% titanium carbide or other alloying elements added, in which case they are commonly called cermets. Ceramics and cermets have a higher hot hardness and are more chemically inert at high temperatures than cemented carbides, making them more wear resistant and permitting higher cutting velocities. The low coefficient of thermal conductivity for these materials causes the heat flow to be primarily directed into the chips, thus reducing the proportion of heat flow into the cutting tool. These materials are more brittle and cannot be used in applications that generate shock loads or cyclic loads induced by the cutting action or thermal cycles. For some metals such as titanium and certain aluminum alloys, a built-up edge and related wear mechanisms readily occur, making these applications unacceptable for ceramic and cermet tooling. These inserts are mounted in tool holders like carbide inserts but are sometimes made thicker in order to ensure adequate strength. This size difference could make it impossible for them to be fit into the same toolholder that is used for cemented carbides.

Ultrahard Tool Materials. Industrial grade single-crystal diamonds, polycrystalline diamond (PCD), and polycrystalline cubic boron nitride (CBN) are finding greater use in tools used for precise, smoother finishes. These very hard materials are very resistant to abrasive wear and have smooth cutting surfaces that reduce the surface shear stress and related friction forces. These smooth surfaces also help prevent built-up edge. Tooling made from these materials can be used on rigid machine tools at very high cutting velocities to produce precision machined surfaces. Both diamond and boron nitride have a cubic structure, but diamond will react with ferrous, materials at moderate temperatures; therefore, diamond tooling can be used only to cut nonferrous and nonmetallic materials. Steels and cast irons can be machined with CBN.

The polycrystalline forms are made by sintering fine single crystals of diamond or CBN together. Lapping is used to produce the geometry for single-crystal tools. Polycrystalline materials can be sintered into blanks or indexable inserts. Despite the costs, the use of ultrahard tool materials can be justified where precise, high-quality surfaces are needed or when higher productivity and fewer tool changes cycles result in competitive operating costs.

Cutting Tool Testing

Testing is done by tool users to evaluate tool vender products, to obtain information for tool replacement, and to establish the machinability of a material. In each of these instances, the amount of wear can be established by direct measurement of the size of the wear surfaces and the amount of damage to these surfaces due to chipping and fracturing. Figure 14 shows the types of wear surfaces that can occur. Such direct measurements require that the cutting process be stopped and optical inspection be used to clearly show the sizes of the wear scars and presence of chipping/fracturing. The deterioration of the tool due to wear can be indirectly established by monitoring surface finish, dimensions, power levels, tool loads, or acoustic events (Ref 16, 17, 18, 19, 20, 21). Correlation studies must be made to relate wear conditions to the indirect measured value.

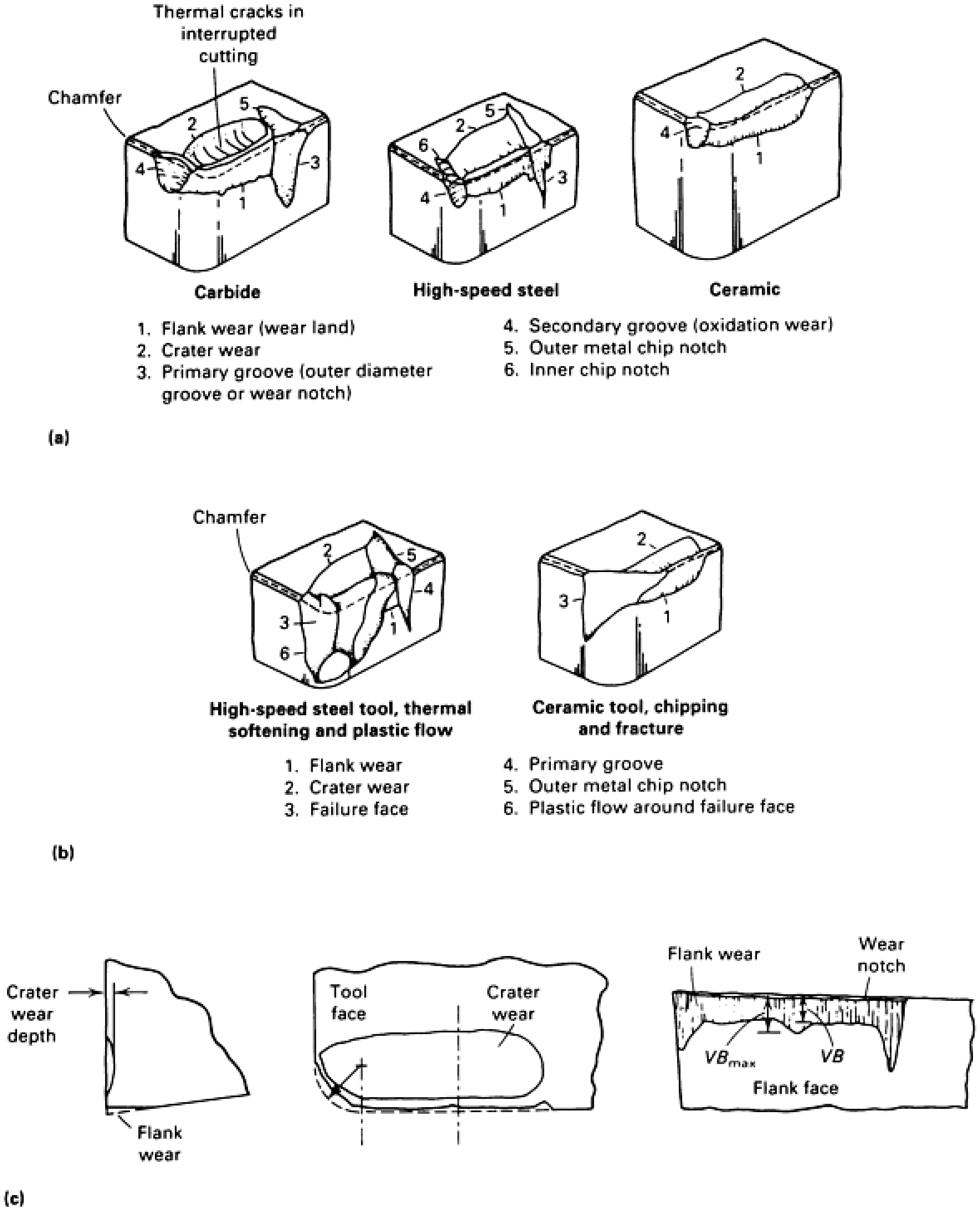


Fig. 14 Cutting tool failure modes. (a) Characteristic wear and fracture surfaces on cutting tools. (b) Catastrophic failure. (c) Typical wear measurements for a turning tool. *VB,* flank wear. Source: Ref 15

Tool life is an important result from testing programs. To establish tool life, a tool failure criterion is needed. This criterion will depend on the cutting application. For example, if the critical wear level is tied to surface finish, then a critical surface roughness level becomes the tool life criterion. To more completely define tool life, the elapsed interval before the critical wear condition is reached should be established for different cutting conditions. Figure 9 shows how a wear measure may vary when one cutting condition such as cutting velocity is varied. The elapsed interval before the wear value reaches the critical value, W_1, varies. This elapsed interval would be the tool life for each of these cutting conditions. The elapsed interval can be time, as in Fig. 9, or the number of parts. Relating this elapsed interval to cutting

conditions produces tool life equations that have been extensively studied and used to establish operating conditions and tool change intervals (Ref 6, 8, 9, 10).

Machinability testing evaluates the cutting characteristics of a material. These characteristics may vary depending upon tool type and tool material. Tool life and the parameters in a tool life equation may be used as one set of measures of machinability; however, they are generally not exclusively used. A general definition of machinability is "the relative ease or difficulty of removing metal in transforming a raw material into a finished product" (Ref 14). Such a comprehensive measure requires information from tool life or wear tests, surface finish tests, cutting force tests, power consumption tests, and cutting temperature tests. Even with reliable results from such a range of tests, the user of these data must combine the information in a way that will provide meaningful comparisons between different materials being machined using different types of tools and tool materials. Methods for doing this are available, but the reader will find that there is not uniform agreement concerning a best measure. The machinability of a material and the performance of particular tool materials in machining a surface can vary considerably depending upon whether or not a cutting fluid is used and, if used, how the cutting fluid is applied.

Cutting Fluid Systems

A cutting fluid can be in the form of a solution or an emulsion. A solution has a base fluid and additives that are soluble in the base fluid. An emulsion uses water as a continuous phase with a discontinuous phase of small suspended particles. Cutting fluids are generally used to improve tool life and increase productivity by allowing higher metal removal rates and decreasing the number of interruptions due to tool failure or replacement. This is generally achieved by lubricating mechanisms that reduce the quantity of heat generated by the cutting process and by cooling mechanisms that control the temperatures in the cutting region.

Solutions of oil or synthetic liquids have the highest lubricating capabilities and the lowest cooling efficiencies. Water-based solutions are better coolants than lubricants. Emulsions often can be formulated to provide both lubrication and cooling.

Cutting fluids often simultaneously control built-up edge on the tool, flush away chips, and protect the workpiece, tooling, and machine from corrosion. With such important purposes, the success or failure of machining operations may depend upon the selection of the cutting fluid system.

The cutting fluid system consist of the cutting fluid, the application mechanism, and the cutting fluid handling system. All three elements of this system must be carefully considered if cooling, lubricating, chip removal, and corrosion control are going to be achieved and maintained. These system issues are best addressed by having a clear goal concerning what the cutting fluid is to accomplish. A related system problem is that this goal must be achieved without endangering the worker and the physical environment. To continuously achieve this goal without environmental damage, the system must control the quality of the cutting fluid. Because of the system issues and varied goals, cutting fluid specialist need to be consulted.

Cutting Fluids. Many different types of cutting fluids are available, and the reader will find extensive literature and vendor help regarding the appropriate type of cutting fluid to use (Ref 22, 23). During such investigations and consultations, the information will be easier to evaluate if there is some basic understanding of the cooling, lubricating, corrosion, and chip removal mechanisms.

Cooling Mechanisms. To cool, the fluid must gain access to the cutting region and must have the thermal capacity to remove the heat generated by the cutting action. The high cutting pressures result in the contact areas between the work and the wear surface that are nearly equal to the projected geometric areas. The lack of void space reduces the amount of cutting fluid that can access the interface region via simple flow mechanisms. Access to these areas is thought to occur via capillary action (Ref 22). To access contacting surfaces, some have suggested that the fluids can penetrate through the metal lattice via diffusion (Ref 24); others have suggested that the cutting fluid may volatilize and the gas phase of the fluid penetrates into the interface (Ref 22). Some form of access must exist if the cutting fluid is expected to reduce friction. The importance of accessibility will vary in that large cuts result in larger contacting wear surfaces. In such instances, friction reduction and related heat generation may have to be controlled by altering other machining conditions, such as reducing cutting velocity.

The cutting fluid thermal properties that are important in removing heat are thermal conductivity, specific heat, heat of vaporization, and wettability with metal surfaces. Emulsions and solutions with greater water contents will have the best

thermal properties. The trend toward higher cutting velocities and larger material removal, rates that inhibit accessibility puts more importance on heat removal. The additives that are present in these solutions and emulsions are selected for their ability to provide some lubrication under such difficulty circumstances. Such lubrication is particularly needed when machining high-quality surfaces.

Lubricating Mechanism. Since a liquid film is unlikely to exist continuously, if at all, in the wear surface contact areas, additives that provide lubrication need an additional mechanism in order to be effective. Additives that promote boundary lubrication are thought to be effective if they react chemically with the workpiece material and tool material to form compounds on the surfaces. These compounds are then believed to interact with the work material to reduce heat generation caused by the strain processes in the work material adjacent to the contacting surfaces (Ref 24). This may or may not be the mechanism, but there is sufficient experimental work to show that these additives do reduce cutting forces, tool wear, and improve surface finish. The fact that some of these benefits lessen at higher cutting velocities suggests that insufficient reaction time is available for compound formation and accessibility is inhibited.

The additives that are effective vary with the severity of the cutting operation. For light cuts, fatty acids can be effective because they are able to form quasi-solid film compounds that are stable up to about 95 to 150 °C (200 to 300 °F). For more severe cutting conditions, extreme-pressure additives are available that form true solid compounds which are stable up to 980 °C (1800 °F).

Corrosion Protection Mechanism. Since fluids that chemically react to corrode metals are needed in cutting fluids to make them effective as lubricants and coolants, other additives are needed to counteract these reactions. Some additives are used to control the pH and obtain an alkaline cutting fluid. Other additives provide a physical barrier by coating the surface of the exposed metal. Establishing and maintaining the correct chemistry is essential. Cutting fluid systems should use treated water of known composition. Careful addition of additives plus the reprocessing of used cutting fluid will maintain the chemistry and prevent biological contamination. Commercial systems are available that provide these control features.

Chip Removal Mechanisms. The need for this mechanism varies with the type of cutting tool and operation. Drilling (deep-hole drilling in particular) depends upon the flow of the cutting fluid to prevent the chips from clogging the flutes of the drill. Other tools that are less confined by the machined surfaces use the flow of the cutting fluid to prevent the chips from remaining in the cutting region where they could interfere with the cutting action and damage the machined surface. If the cutting fluid has a tendency to foam when severely agitated, the flow rates and flow patterns needed for chip removal may need to be carefully controlled because the foam can clog filters in the circulation system.

Application Methods. Cutting fluids that are primarily used to lubricate must be applied so that they form films on sliding surfaces, whereas cooling applications require good contact with the cutting region tool surfaces. The need to remove chips imposes another application requirement. Obviously, the application technique will be closely tied to the goal selected for the cutting fluid system.

Hand application of cutting fluids is used only in low-production situations because it is difficult to continuously apply sufficient quantities of cutting fluid by hand to cool the tool. When used for drilling and tapping, it is best to hand apply the cutting fluid periodically when the drill or tap is backed out of the hole to clear chips.

Flooding is the most common application method because it is effective at cooling, can provide good lubrication, and is an effective way to remove chips. The volume required depends upon the cutting operation, the size of the surface feature being machined, and the cutting conditions (tool engagement, feed, and cutting speed). The number of nozzles and their orientation is as important as the quantity of cutting fluid used in flooding. The cutting region should initially receive the cutting fluid; in some instances, a particular portion of the cutting region will receive more fluid than another region. In turning and boring, the primary application of fluid should be directed at the chip and cutting edge region. A secondary application area is the clearance surfaces where lubricating films reduce flank wear, particularly at slower cutting speeds. For drilling, the task is to get the cutting fluid down the hole and the cutting chips up out of the hole. Proper alignment of the nozzles and proper volumes and pressures can be effective, but more drilling tools now have cutting fluid holes positioned through the shank so that the fluid is delivered directly to the cutting edge region. For milling, it is desirable to have at least two nozzles--one directed to the cutting region and the other used to clear the chips.

Mist application is possible where smaller engagements and high cutting speeds are used. Mist particles easily vaporize upon contact, efficiently removing heat during the vaporization process and reducing the quantity of cutting fluid needed

for the application. Mist application improves visibility of the cutting region. A danger in misting is that, if not properly controlled, suspended particles in the air will contaminate the working environment.

Very-High-Pressure Application. Recently, several techniques have been developed to provide very-high-pressure coolant streams for cooling lubrication and chip control. Most of these techniques are in the development stage, but a few are beginning to emerge as products. These very-high-pressure streams appear to reduce the chip pressures against the rake surface and help the coolant gain access to the wear surfaces (Ref 25).

Cutting fluid systems can range from simple hand application to plant-wide distribution systems served from a common recirculation and cleaning system. It is very important that the system protect the workers from exposure to cutting fluids that pose health risks. The system should maintain a level of cleanliness that is consistent with accepted health standards for the industry. The cleaning of the recirculated cutting fluid is very important. Tramp oil from machinery lubricating systems and metal chips must be removed as well as much finer contaminants. Contaminants can come from the cutting process or from the plumbing of the circulation system, they can even exist as microbes that biologically grow in such environments. All must be removed and the chemistry adjusted back to the specified level. Lastly, and most importantly, the system must safely dispose of its waste materials.

References

1. J.T. Black, Mechanics of Chip Formation, *Machining,* 9th ed., Vol 16, *Metals Handbook,* ASM International, 1989, p 7-12
2. M.C. Shaw, *Metal Cutting Principles,* Oxford University Press, 1984
3. S.C. Lim and M.F. Ashby, Overview No. 55, Wear-Mechanism Maps, *Acta Metall.,* Vol 35 (No. 1), 1987, p 1-24
4. P.K. Wright, Physical Models of Tool Wear for Adaptive Control in Flexible Machining Cells, *Proceedings of the Symposium on Computer Integrated Manufacturing,* Vol 8, American Society of Mechanical Engineers, 1983
5. B.M. Kramer and N.P. Suh, Tool Wear by Solution: A Quantitative Understanding, *J. Eng. Ind. (Trans. ASME),* Vol 102, 1980, p 303
6. F.W. Taylor, On the Art of Cutting Tools, *Trans. ASME,* Vol 28, 1907
7. B.N. Colding and W. Konig, Validity of the Taylor Equation in Metal Cutting, *Ann. CIRP,* Vol 19 (No. 4), 1971, p 793
8. V.A. Tipnis, Cutting Tool Wear, *Wear Control Handbook,* M.B. Peterson and W.O. Winer, Ed., American Society of Mechanical Engineers, 1980, p 891-930
9. S.M. Wu, Tool Life Testing by Response Surface, Methodology, Parts I and II, *J. Eng. Ind. (Trans. ASME),* Series B, Vol 86, 1964, p 105
10. A.K. Sheikh, L.A. Kendall, and S.M. Pandit, Probabilistic Optimization of Multitool Operations, *J. Eng. Ind. (Trans. ASME),* Vol 102, 1980
11. K.S.A. Brooks, *World Directory and Handbook of Hard Materials,* 4th ed., Engineers Digest and International Carbide Data, 1987
12. *1985 Standards on Cutting Tool Wear,* American National Standards Institute and American Society of Mechanical Engineers
13. T.J. Drozda and C. Wick, Ed., *Cutting Tool Materials,* Society of Manufacturing Engineers, 1988
14. C. Zimmerman, S.P. Boppana, and K. Katbi, Machinability Test Methods, *Machining,* 9th ed., Vol 16, *Metals Handbook,* ASM International, 1989, p 639-647
15. S. Kalpakjian, *Manufacturing Processes for Engineering Materials,* Addison-Wesley, 1984, p 489
16. S. Jetly, Measuring Cutting Tool Wear On-Line: Some Practical Considerations, *Manuf. Eng.,* July 1984, p 55
17. E.N. Diei and D.A. Dornfeld, Acoustic Emission Sensing of Tool Wear in Peripheral Milling, *J. Eng. Ind. (Trans. ASME),* Vol 105 (No. 3), 1987, p 234-240
18. J.E. Mayer, Cutting Tool Monitoring, *COMMLINE,* Nov-Dec 1985

19. K.W. Yee and L. Evans, Drill-Up, An Alternative for On-Line Determination of End-Mill Wear, *13th North American Manufacturing Research Conference Proceedings,* Society of Manufacturing Engineers, May 1985
20. A.E. Bayoumi and L.A. Kendall, Modeling and Measurement of Wear of Coated and Uncoated High Speed End Mills, *J. Mater. Shaping Technol.,* Vol 16 (No. 1), 1988
21. A. Thangaraj and P.K. Wright, Computer-Assisted Prediction of Drill Failure Using In-Process Measurement of Thrust Force, *J. Eng. Ind. (Trans. ASME),* Vol 110, 1988, p 192-200
22. T.J. Drozda and C. Wick, Ed., *Cutting Fluids and Industrial Lubricants,* Society of Manufacturing Engineers, 1988
23. E.S. Nachtman, Metal Cutting and Grinding Fluids, *Machining,* 9th ed., Vol 16, *Metals Handbook,* ASM International, 1989, p 121-131
24. J.G. Horne, D. Tabor, and J.A. Williams, Action of Gaseous and Liquid Lubricants, on Metal Cutting, *Proceedings of the International Symposium on Metalworking Lubrication,* American Society of Mechanical Engineers, 1980 p 193
25. M. Mazurkiewicz, Z. Kubala, and J. Chow, Metal Machining with High-Pressure Water-Jet Cooling Assistance--A New Possibility, *J. Eng. Ind. (Trans. ASME),* Vol III, 1989

Selected References

- A.M. Bayer and B.A. Becherer, High-Speed Tool Steels, *Machining,* 9th ed., Vol 16, *Metals Handbook,* ASM International, 1989, p 51-59
- Cast Cobalt Alloys, *Machining,* 9th ed., Vol 16, *Metals Handbook,* ASM International, 1989, p 69, 70
- W.W. Gruss, Cermets, *Machining,* 9th ed., Vol 16, *Metals Handbook,* ASM International, 1989, p 90-97
- P.J. Health, Ultrahard Tool Materials, *Machining,* 9th ed., Vol 16, *Metals Handbook,* ASM International, 1989, p 105-117
- R. Komanduri and S.K. Samanta, Ceramics, *Machining,* 9th ed., Vol 16, *Metals Handbook,* ASM International, 1989, p 98-104
- K.E. Pinnow and W. Stasko, P/M High-Speed Tool Steels, *Machining,* 9th ed., Vol 16, *Metals Handbook,* ASM International, 1989, p 60-68
- A.T. Santhanam and P. Tierney, Cemented Carbides, *Machining,* 9th ed., Vol 16, *Metals Handbook,* ASM International, 1989, p 71-89

Friction and Wear of Dies and Die Materials*

Rajiv Shivpuri, Department of Industrial Engineering, Ohio State University; S. Lee Semiatin, Materials Directorate, Wright Laboratory

Introduction

WEAR OF DIES AND MOLDS is a complex time-dependent phenomenon that primarily depends on the four components of the system: die (mold), interface, workpiece, and processing conditions. The effect of these four components can be categorized into six process-related considerations: die (mold) design, die (mold) material, heat treatment, lubrication, surface treatments and coatings, and processing conditions.

Dies or molds impose a geometry on the deforming material. The design of the die cavities governs the sliding velocities, temperature, and interface pressures. Especially important considerations from a wear and failure standpoint are the radii of corners and fillets in die cavities. For example, in die casting the design issues are: size of die cavity with respect to the die blank; distance of cavity from the gates, cooling lines, and outer surface; cavity contours (radius of stress intensifiers); and sudden changes in cavity cross sections.

Another important consideration for wear is the micro- and macrostructural properties of the die material, its composition, microstructure (uniformity, internal defects, and secondary carbides), and mechanical and physical properties. Mechanical and chemical interaction of the die material with the rubbing body (workpiece) and the contaminants (lubrication, debris, and scale) is dependent on composition and microstructure.

Heat treatment of dies has a major influence on the resultant microstructure, hardness, and toughness of the dies. The casting cycle or quenching cycle determines the amount of pearlite, bainite, and martensite microstructures, which control the hardness, toughness, and temper resistance. Decarburization often results and has to be accounted for. Also of considerable importance is the austenizing temperature of the treatment.

Lubrication is important in that it determines the frictional conditions at the interface. In addition, it alters the contact conductance (thermal coefficients) and the oxidation behavior of the interface. In die casting and injection molding, lubricants also react with the melt. Experimental studies in hot forming suggest that sometimes good lubrication enhances abrasive wear. This is because lower interface friction permits larger sliding velocities at the interface and, consequently, higher wear rates. In addition, lubrication inhibits the formation of brittle (untempered martensite) layers on the die surface.

Surface treatments and coatings have found increase usage for wear resistance. Some of the important considerations are coating thickness, thermal and chemical stability, adhesive strength, spalling behavior, and cost. Exotic coatings such as ceramic coatings have found use in high-temperature applications, such as hot extrusion, and in drawing dies.

Processing conditions such as mechanical and thermal loading histories and sliding distances and velocities have a profound effect on die wear behavior. Factors such as maximum temperature, die preheating temperature, and dwell time (contact time under load) are critical to the success of hot-forming operations. Acceptable die lives can be obtained by proper control of processing factors. Forming equipment influences the wear process by imposing predetermined mechanical time histories on the process.

Acknowledgements

This work was possible due to the support extended by the National Science Foundation-sponsored Engineering Research Center for Net Shape Manufacturing at the Ohio State University. Special thanks are due to Professor Taylan Altan, Director of the Center, for his support and encouragement.

Note

* *Adapted from "Wear of Dies and Molds in Net Shape Manufacturing," Report ERC/NSM-88-05, National Science Foundation Engineering Research Center for Net Shape Manufacturing, Ohio State University, 1988

Die Wear and Failure Mechanisms

Wear of materials occurs by many different mechanisms. The terminology used to describe these mechanisms depends on the field of application. Some of the wear mechanisms identified in die and mold wear and failure are:

- Adhesive wear
- Abrasive wear
- Erosive wear
- Corrosive wear
- Thermal fatigue (heat checking)

- Mechanical fatigue
- Thermal shock
- Plastic deformation
- Gross cracking

The first three of the aforementioned wear mechanisms are common to most forming processes.

Adhesive wear is the predominant wear mechanism in the cold adhesion and galling behavior in sheet metal forming and cold extrusion. Surface finish and hardness play an important role in adhesive wear, with similar materials (rubbing pair) showing greater wear than dissimilar materials. Surface treatments such as nitriding are preferred in adhesive wear applications.

Abrasive wear is the predominant wear mechanism in hot forging. The surface films such as oxide scale break down during the deformation process and act as abrading particles. This three-body wear situation is also present in other hot-forming processes. In the manufacturing of composites, however, the second phase (reinforcement) acts as the abrasive. Sliding distance and velocity, and interface pressures are important considerations. The hardness ratio (hardness of the abrasive/hardness of the surface) has a major influence on the amount of wear. Surface platings and coatings that increase surface hardness significantly reduce abrasive wear. Care must be taken in the use of coatings because they can crack in service and lead to enhanced wear rates due to spalling.

Erosive wear is predominant in solidification processes such as die casting. The liquid metal spray impinges on the die surface resulting in localized loss of material due to erosion. Temperature differential between the molten metal and the die surface, the velocity of the molten spray, and the angle of impact are some of the important factors in die erosion.

Corrosive wear occurs in combination with the other wear mechanisms described above. It results from the chemical affinity between the flowing material and the dies. Grain size and the composition of the die material influence corrosive wear.

Thermal Fatigue. The most common reason for discarding die casting dies is the development of heat checks (craze cracks) due to thermal fatigue. This also arises during hot forging due to the temperature difference between the dies and the hot metal. As the die surfaces are heated, compressive stresses develop due to constraints provided by the cooler interior material. This situation is reversed when the part is unloaded and the die surfaces cool. Cyclic compressive and tensile stresses during production result in thermal fatigue. The most important factor in thermal fatigue is the strain amplitude imposed during the forming cycle. In turn, this depends on the magnitude of temperature change during each cycle and the thermal properties of the die material. Other factors include composition and purity of the die material. Composition influences the metallurgical phase transformations, which lead to additional large increments of straining above that which leads to failure. The presence of second-phase particles and other inclusions decreases the hot ductility of the material considerably.

Mechanical fatigue of forming dies is influenced by the magnitude of applied loads, the average temperature of the dies, and the surface condition of the dies. Fatigue cracks initiate at locations of stress concentrations (cavities, corners, and fillets) or features such as hotels, keyways, and deep stamp markings on die sets. Design changes to minimize stress concentrations are often recommended. Surface treatments, including nitriding, micro-shotpeening, and ion implantation, effectively reduce surface fatigue. However, plating and coating treatments may lead to a brittle layer that cracks under cyclic loading and causes spalling.

Thermal shock in die casting often causes heat checking and occasionally, gross cracking. The temperature difference between the die surfaces and the liquid metal is the driving factor. Important material properties for thermal shock are the heat capacity, conductivity, and expansion coefficient. Elastic modulus of the material is also of significance.

Plastic deformation in dies results from excessive pressure and low hot-yield strength of the die material; the former can be alleviated by cavity redesign, and the latter by proper selection of the die material. Die cooling also reduces these effects, but care must be taken to avoid temperature cycles, which may cause phase transformations.

Gross cracking, or catastrophic die failure, is a limiting condition from a die life standpoint. Failure occurs quickly (in a few cycles) because of high applied stress or low toughness of the die material. As in the case of mechanical fatigue,

high stresses may result from poor die design, improper press fitting or shrink fitting of dies and die inserts, or lack of control of forging load and energy.

Materials for Dies and Molds

A wide range of materials are available to the designer of tools and dies. This section summarizes the important attributes required of dies and the properties of the various materials that make them suitable for particular applications. Among the important attributes are hardenability; machinability; and resistance to wear, plastic deformation, shock loading, and heat checking. The needed levels of resistance to wear, plastic deformation, and so forth, are determined by factors such as type of equipment used, workpiece temperature, expected die temperature, and number of parts to be fabricated. Low-alloy steels and hot-work die steels are often suitable for conventional metalworking, casting, and molding operations. On the other hand, high-temperature die materials are required for special applications such as isothermal forging of titanium and nickel-base alloys. These die materials include various superalloys and TZM molybdenum. Recommendations on the selection of these materials are made in this section. The approach used in making these recommendations and the tool materials themselves are discussed in detail for hot forging tooling. Short summaries of material recommendations are also provided for other fabrication techniques.

Die Materials for Hot Forging

The selection of materials for hot-forging dies is done by considering the types of loads and temperatures that are to be imposed on the dies. Once this is known, an evaluation of die material properties can be made with respect to the known loads and thermal conditions that the dies will experience. The important factors in the selection procedure and their interrelation are schematically shown in Fig. 1. The important forging process variables include the following:

- Type of forging equipment employed (which determines the maximum load that can be applied and loading rate)
- Workpiece material and its preheat temperature (which affect die pressure and temperature)
- Size and shape of the part (which influence overall die loading)
- Type of lubrication and die heating/cooling system (which affect the thermal history and surface wear of the dies)
- The number of parts to be forged (which determines the number of load and thermal cycles)
- The production rate

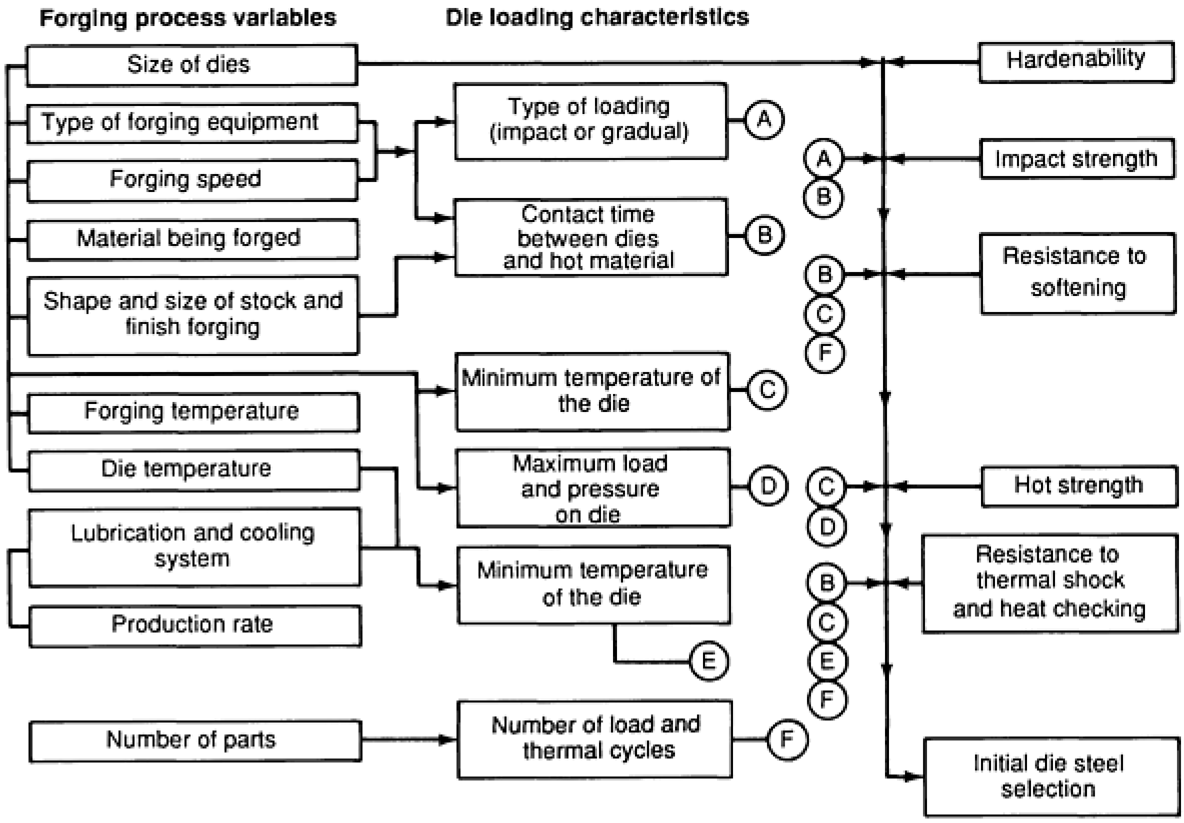


Fig. 1 Schematic representation of die steel selection procedure. Source: Ref 1

In turn, the die material attributes of importance are hardenability; machinability; wear resistance; hot strength; toughness; resistance to heat checking, thermal fatigue, and mechanical fatigue; possible need of a special protective atmosphere in the forging operation; and the material cost (Ref 1, 2).

Alloy Steels for Hot-Forging Dies. A variety of die steels are available for warm and hot forging (Ref 1, 3). These steels have varying degrees of hot strength, toughness, and fatigue resistance. Before comparing specific steels, however, it is instructive to review the compositions and attributes of the various grades.

Low-Alloy Tool Steels (6G, 6F, 6H). The principal alloying elements in these steels are nickel, molybdenum, and chromium, with vanadium and silicon in smaller additions. The total alloying content is generally small enough so that adequate machinability is retained in prehardened die blocks of these grades. Characterized by high toughness and, in some instances, by good heat resistance (6F4), these steels have very good hardenability. As a tradeoff to their good toughness, they are generally heat treated to relatively low hardnesses. Because of these low hardness, their wear resistance is only moderate. However, they are known to possess good resistance to shock loading (such as encountered in hammer forging), heat checking, and catastrophic failure.

Because of the generally low tempering temperatures, these die steels are employed primarily in hammer operations where the contact times, during which heat transfer to the dies can occur, are short. Alternatively, they can be used in presses as die blocks with heat-resistant inserts made from the hot-work die steels (H grades). An exception to this general trend among the low-alloy tool steels is alloy 6F4. Its large nickel and molybdenum contents impart high hardness and hardenability. Furthermore, it is generally underaged. Because of this, when it is used in presses or upsetters, heat transfer gives rise to age hardening and thus increased resistance to abrasive wear.

Air-Hardening Medium-Alloy Tool Steels (A2, A7, A8, A9). Manganese, chromium, molybdenum, and vanadium are the principal alloying elements in this group of tool steels. These steels have moderate resistance to thermal softening, and, because of their high carbon content, have high strength and good wear resistance. Hence, they are useful for both hot and cold forging.

Chromium Hot-Work Steels (H10 to H19). These steels contain chromium as the major alloying element with additions of molybdenum, tungsten, vanadium, and cobalt. They possess good resistance to thermal softening and heat checking, and moderate toughness at working hardnesses of 40 to 55 HRC. In addition, the tungsten and molybdenum promote good hot hardness. The alloy and carbon contents of these steels are low enough so that they can be water cooled usually without cracking and consequently have improved service life. Also, because of their high hardenability, they can be air hardened (air-cooled after austenitizing or tempering), which minimizes distortion after heat treatment.

As for specific steels in this group, H10 has high resistance to thermal softening. Grade H11 also has high resistance to elevated-temperature softening, but it offers improved toughness over H10. Furthermore, the high tempering temperature of H11 permits relief of residual stresses, thus leading to an optimal blend of strength and toughness. In grade H12, tungsten is added to improve the toughness, hot hardness, and heat checking characteristics over those of H11. However, the presence of tungsten decreases resistance to thermal shock; therefore, water cooling of H12 is not recommended. In H13, a high vanadium content increases resistance to heat checking (below 650 °C, or 1200 °F) and abrasive wear. Also, H13 appears to have good thermal shock resistance and can be water cooled if desired. Grade H19 with its large tungsten and cobalt additions and moderate vanadium addition has excellent hot hardness and resistance to abrasive wear. The presence of tungsten, however, decreases the alloy's resistance to thermal shock; therefore, like H12, H19 should not be water cooled. Because of its high cost and only moderate toughness, H19 is typically used only for inserts in forging dies.

Tungsten Hot-Work Steels (H20 to H26). The principal alloying element in these die steels is tungsten. Various grades contain additions of chromium and vanadium. The higher alloy content of these steels over the standard chromium hot-work steels leads to improved resistance to thermal softening and abrasive wear, but at the same time it gives them only moderate to low toughness at their normal working hardnesses of 45 to 55 HRC. In addition, the high alloy content makes this group of steels susceptible to thermal shock; therefore, they should not be water cooled. If these steels are preheated to operating temperature before use, they can be used for forging dies of rugged design.

Molybdenum Hot-Work Steels (H41 to H43). Molybdenum is the principal alloying element in this group, whose grades also contain varying amounts of chromium, vanadium, and tungsten. The characteristics of these steels are very similar to those of the tungsten hot-work steels, the major difference being the lower cost of the molybdenum hot-work steels. Another advantage of these steels over the tungsten hot-work steels is their greater resistance to heat checking, but, as with all high-molybdenum steels, they must be heat treated carefully to avoid decarburization.

Comparison of Die Steels for Hot Forging. Selection of a die steel for forging depends on careful consideration of the die material's metallurgy as well as its mechanical and thermal properties. Metallurgical factors include hardenability and machinability. Mechanical properties include the following:

- *Resistance to Wear.* The ability to resist the abrasive action of the workpiece metal sliding over the dies during forging. This property is affected by the hot hardness of the die material and its resistance to thermal softening
- *Resistance to Thermal Softening.* The ability of the die material to resist softening, or overaging, during long-term elevated-temperature exposure
- *Resistance to Plastic Deformation.* The ability of the die material to resist deformation under forging loads and expected service temperatures during forging (Fig. 2 and 3)
- *Resistance to Brittle Fracture.* The ability to resist catastrophic failure and is determined by the alloy toughness
- *Resistance to shock loading* as may be experienced in hammers
- *Resistance to Heat Checking.* The ability to withstand surface cracking due to nonuniform die heating (dies with a hot surface and cooler interior)

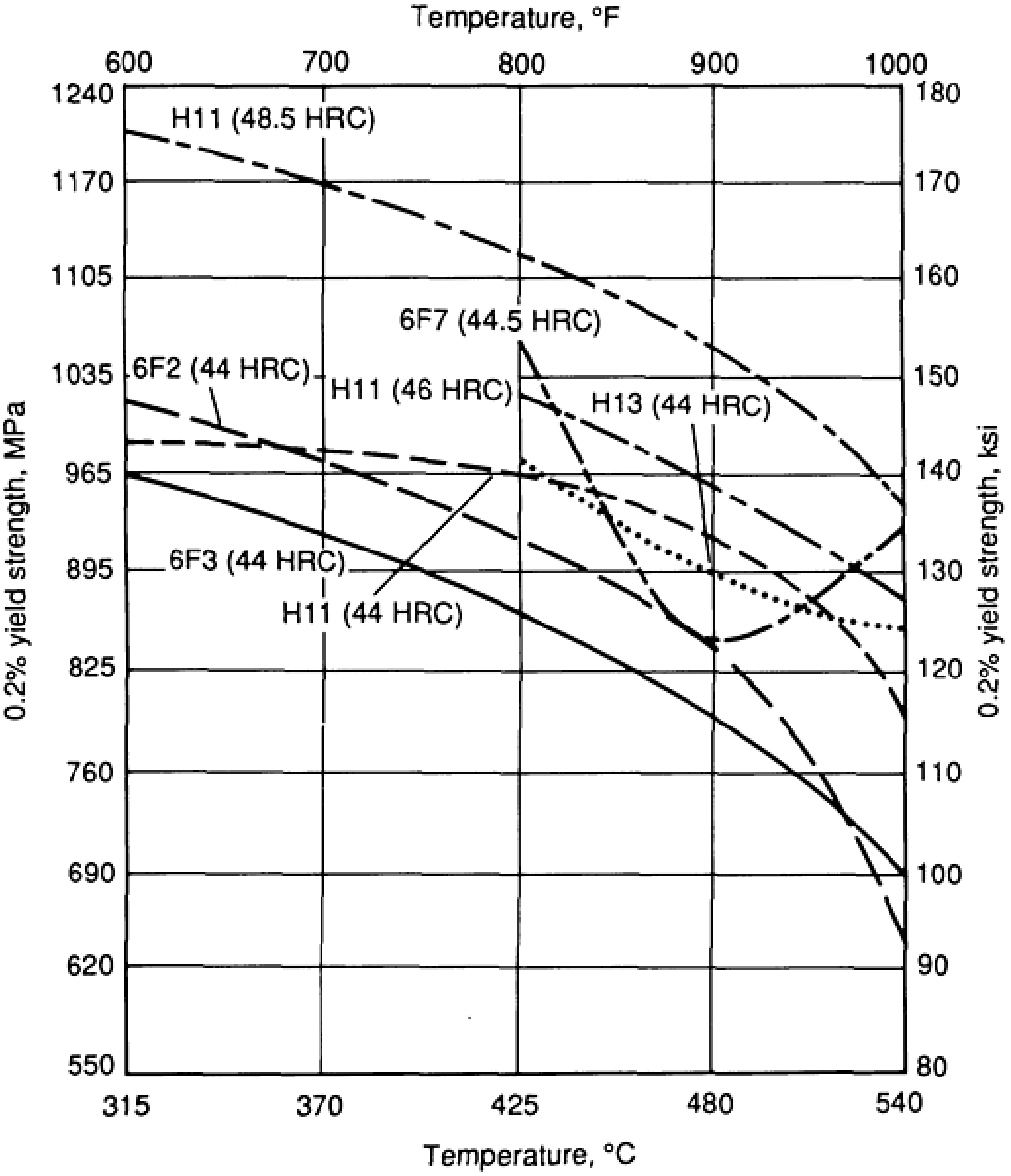


Fig. 2 Resistance of selected die steels to plastic deformation at elevated temperatures. Values in parentheses indicate room-temperature hardness. Source: Ref 4, 5

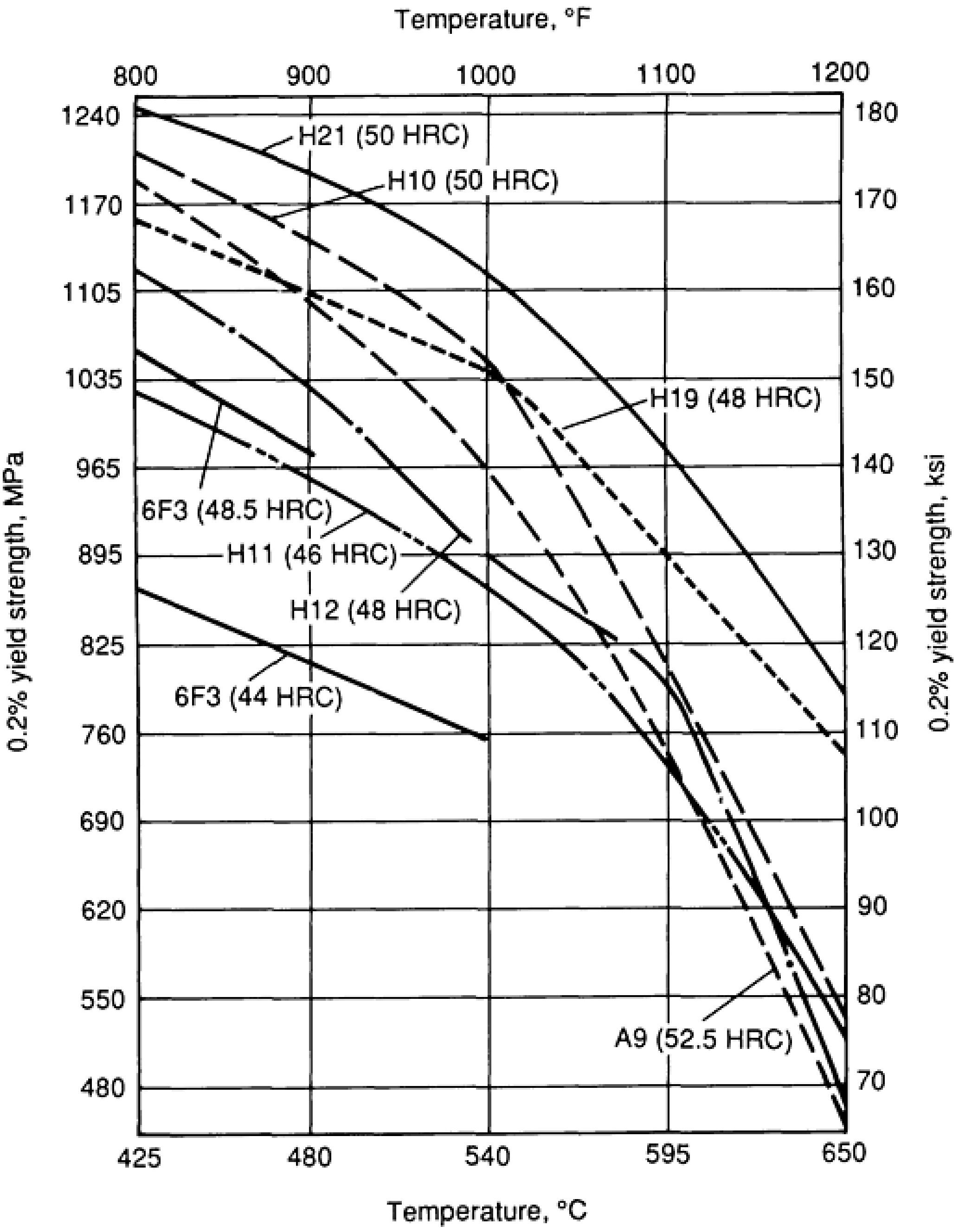


Fig. 3 Resistance of selected die steels to plastic deformation at elevated temperatures. Values in parentheses indicate room-temperature hardness. Source: Ref 4, 5, 6

Rating of Die Steels for Hot Forging. Based on the previous discussion and additional literature supplied by steel manufacturers, die steels can be qualitatively ranked for relative response to wear, impact, high temperature, and so forth. Such a rating, of course, is helpful in selecting among the many materials currently available. Table 1 presents a relative ranking for three of the most important die material properties. The rating on heating-checking quality is not provided because conclusive data are not available. It must be emphasized that the comparative rating is arbitrary. This is because other factors, such as hardness of the die block and the maximum temperature to which the dies are subjected, influence the comparison. For example, H21 hardened to a higher level than H19 may show better wear resistance than H19; in the same context, 6F2 may show a wear resistance comparable to that of H11 if the maximum die temperature does not exceed 370 °C (700 °F). Thus, for a specific application, a closer comparison of the mechanical properties of the die steels is needed to make a selection.

Table 1 Die steel ratings

	Resistance to		
	Wear	Thermal softening	Catastrophic fracture (impact, strength, toughness)
	H26	H26	6F2, 6F3 6F5, 6F7
	H23	H24	
↑	H24, A2	H23	H12
	H19, H14	H19	H11
	H21	H21	H10, H13, 6F4
Increasing resistance	H10, H12, H13	H14	A9, 6H2, 6G
	A9, 6H2, A8	H10	H14
	H11, 6F4	6F4	6H1, A8
	6H1, 6F5 6F7, 6F3	H11, H12, H13, A8	H19, A2
	6F2, 6G	A9	H21, H23, A6
		6F3, 6F5, 6F7, 6H1, 6F2, 6G	H24
			H26

Source: Ref 1, 6–8

Source: Ref 1, 6, 7, 8

High-alloy steels such as H26, H19, H21, and H14, possess excellent wear and heat-resistance qualities, but have poor toughness. Thus, these steels are best used in forging dies as inserts that can be rigidly supported in the die blocks. In hammers where toughness is the primary requirement, these steels are not normally used for die blocks. Low-alloy steels such as 6G, 6F2, and 6F3 show the opposite qualities. These steels have low wear- and heat-resistance qualities, but high toughness. High resistance to impact combined with the low material costs make these steels very attractive for hammer dies. Medium-alloy steels such as H11, H12, and H13 provide a compromise between the two extremes. These steels are especially suitable for press dies where the die steel must have good resistance to thermal softening as well as reasonable toughness. Medium-alloy steels also find application in hammers where the quantities of parts to be forged are large, and wear resistance becomes an important economic requirement in such applications.

In addition to standard steels, special steels, such as modifications of the H10 and H13 grades, are available for applications where high wear resistance is desired and a reduction in toughness can be tolerated. For large die blocks where hardenability and strength throughout the block are added requirements, 6F5 and 6F7 are especially suitable. Thus, depending upon the needs of a particular forging operation, a baseline selection of suitable steels can be made from the data on mechanical properties of the die steels.

Effect of Process Variables and Workpiece Properties. As mentioned previously, forging process variables and workpiece properties must be considered carefully when selecting die steels in order to ensure that the dies can withstand the loads and temperatures to which they are to be subjected. These considerations are briefly discussed below. Also, specific die steel recommendations for particular combinations of workpiece material and equipment will be given.

Characteristics of Forging Equipment. The main types of equipment used in forging are hydraulic presses, mechanical presses, upsetters, screw presses, and forging hammers. In a hydraulic press, the ram moves slowly, giving rise to large amounts of die heating (because of long contact times) and large loads. Thus, the dies must have high hot strength and be resistant to thermal softening and wear. Toughness and resistance to impact are minor considerations. For these reasons, the chromium hot-work steels (H10 through H16) are often employed. Alternatively, the low-alloy steel 6F4, which hardens during service, can be effectively used.

Mechanical presses and upsetters are similar to hydraulic presses in that forging with these types of equipment leads to relatively high loads and long contact times. Again, the die steels must have high hot hardness and resistance to thermal softening and wear, and usually the chromium hot-work steels or an age-hardening alloy such as 6F4 are used. For screw presses, speeds are comparable to mechanical presses, but because the ram is accelerating when it hits the workpiece, die materials that are somewhat more impact resistant (tougher) than those for mechanical presses are required. In Europe, where screw presses are popular, the most commonly used die steel is H10 (Ref 9).

Die steels for forging hammers require different characteristics than those for presses or upsetters. The very rapid loading rate and short contact time in hammers require die steels that are tough, but that have only moderate resistance to thermal softening. Low-alloy steels such as 6F2 are commonly used for die blocks, whereas 6F2, 6F3, or the chromium hot-work steels are employed for die inserts.

Number of Parts to Be Forged. As a minimum requirement for the production of acceptable forgings, whether the lot quantity is 100 or 100,000, the dies must have adequate hot strength at the service temperature to withstand the forging pressure. However, when the number of parts to be forged is very small, or when the design of the forging is likely to be modified, die life is not the determining factor in die material selection. The least expensive prehardened steel that meets the minimum requirement becomes the logical choice for small-lot applications. Higher-alloyed steels are usually more cost-effective for larger lot quantities because they lower the cost of machining and changing die sets.

Size of the Dies. For small dies, the cost of die material represents a minor portion of the total forging cost; the portion of total cost represented by the tool steel costs increases as the size of the die increases. Thus, for small dies, the tendency is toward using higher-strength steels, which may be more expensive in material cost.

For large die blocks, added requirements are hardenability for uniform properties and toughness. Die steels with high nickel content (6F5, 6F7) are, therefore, used for very large die blocks.

Other Process Variables. Forging speed, die temperature, die lubrication and cooling systems, production rate, shape and size of the stock, and finish shape and size of the forging are the other variables that influence the die-loading conditions, as shown in Fig. 1. These parameters influence the maximum die load and maximum die temperature. The lubrication and cooling systems, together with the die temperature, determine the minimum temperature to which the dies are cooled. Maximum and minimum temperatures during a forging cycle define the conditions of thermal shock to which the die is subjected.

Production rate influences the maximum temperature of the dies. With high production, the die life decreases because of the increase in maximum die temperature.

Workpiece Properties. The workpiece properties play a critical role in determining the forging pressure applied to the dies. In addition, the required loads as well as workability of the material are strong functions of temperature. Thus, the workpiece material strongly affects the die temperatures that can be expected in forging production, particularly for presses and upsetters.

Forging of carbon steels is easy and requires relatively inexpensive die block materials. Typically, low-alloy tool steels are sufficient (Table 2). When forging is done in presses, the dies are usually heat treated to a higher hardness than they would be if forging were done in hammers because wear resistance is more important and toughness less important. In some instances, though, die inserts of the more highly alloyed low-alloy steel, or even of the chromium hot-work die steels, are recommended in regions of the dies exposed to higher-than-average-temperatures or loads.

Table 2 Die steels for forging various alloys

Material	Hammer forging, die steel/hardness, HRC	Press forging die steel/hardness, HRC
Carbon steel	6G, 6F2/37-46 (Die blocks)	6F3, H12/40-46 (Die blocks)
	6F3, H12/40-48 (Inserts)	H12/42-46 (Inserts)
Alloy and stainless steel	6G, 6F2/37-46 H11, H12, H13/40-47 (Die blocks)	H11, H12, H13/47-55 (Die blocks, inserts)
	H11, H12, H13, H26/40-47 (Inserts)	
Aluminum alloys	6G, 6F2/32-40 (Die Blocks)	6G, 6F2, 6F3/37-44, H12/47-50 (Die blocks)
	H11, H12/44-48 (Inserts)	H12/46-48 (Inserts)
Titanium alloys	6G, 6F2/37-40 (Die blocks)	6G, 6F2/37-40 (Die blocks)
	H11, H12/44-52 (Inserts)	H11, H12/47-55 (Inserts)
Heat-resistant alloys, nickel-base alloys	H11, H12, H13/47-50 (Die blocks, inserts)	H11, H13, H26/50-56 Inconel 713C, René 41 (Die blocks, inserts)

Source: Ref 3

For forging low- to medium-alloy steels and stainless steels, more stringent demands are made on the forging dies and die materials. In hammer forging, die blocks can often be made of low-alloy steels (Table 2). However, small dies or die inserts should be made of hot-work die steels. For press forging these alloys, chromium hot-work steels are often used for both dies and die inserts, with die inserts usually tempered to slightly higher hardness than the die blocks.

Typical forging die materials for aluminum alloys are very similar to those for carbon steels (Table 2). Prehardened low-alloy steel die blocks are often used. For aluminum alloys, though, they are tempered to slightly lower hardnesses than dies for forging carbon steel. As for carbon steels, it is often advantageous to use die inserts of more highly hardened chromium hot-work steels in critically stressed regions of both hammer and press dies.

Titanium alloys are similar to stainless steels in that forging dies for these materials are usually made from chromium hot-work steels (Table 2). For large dies, the die impression is made of one of these steels and inserted into a large die block of a low-alloy steel such as 6G or 6F2. As for hardness, hammer dies are only moderately hardened so that they retain a certain amount of toughness. On the other hand, press dies are usually harder than hammer dies (by about 3 HRC points) because of the reduced need for toughness. However, the low end of the recommended hardness range of press die inserts should be used for dies with severe impressions.

The greatest demand upon die material is made when forging heat-resistant alloys and nickel-base superalloys. This is not surprising inasmuch as they have been designed to resist deformation at high temperatures. The primary choices for dies and die inserts for these alloys are the chromium hot-work steels. As for the forging of other alloys, these steels are typically more highly hardened for press application than for hammer applications. In general, the hardness levels used tend to be quite high, primarily to resist wear caused by very high die pressures. For these kinds of demanding service conditions, the tungsten hot-work die steels, such as H26, or nickel-base alloy die materials, such as Inconel 713LC or IN-100, may offer and advantages when considering die life versus material and machining cost.

Other Die Materials for Hot Forging. Several other kinds of steel exist that may be suitable for forging dies. These steels typically have been designed for specialized applications. Nonferrous materials, such as superalloys, TZM molybdenum, and cemented carbides, are also sometimes used for severe applications. Table 3 compares service temperatures of die materials used in forging operations.

Table 3 Typical service temperature of die materials in forging

Tool material	Recommended service temperature range	
	°C	**°F**
Low-alloy steels, air-hardening steels, shock-resisting steels	205-480	400-900
Chromium, tungsten, and molybdenum hot-work steels, maraging steels, tungsten carbide	370-620	700-1150
Superalloys	620-925	1150-1700
TZM molybdenum	925-1205	1700-2200

In shock-resisting tool steels (AISI S series), the principal alloying elements are silicon, manganese, chromium, tungsten, and sometimes molybdenum or nickel. These elements impart high strength and toughness and moderate wear resistance. Their primary application is hammers in which tool temperatures do not exceed 540 °C (1000 °F).

Maraging steels are heat treated somewhat differently than the conventional die steels, which are quenched and tempered. Containing nominally 18% Ni 8 to 12% Co, and smaller amounts of molybdenum and titanium, these steels from a comparatively soft martensite during air cooling following austenitizing. Subsequent hardening is then performed by aging treatments at approximately 480 to 510 °C (900 to 950 °F). Because these steels distort negligibly during aging, they offer a considerable advantage as far as machining is concerned. Among the properties that make them attractive as compared to conventional hot-work die steels are their significantly higher yield strengths (room-temperature strengths between 1380 and 2410 MPa, or 200 and 350 ksi, depending on the specific alloy) and toughnesses (Ref 10). At the same hardness, their toughness can be at least 2 to 2.5 times that of chromium hot-work tool steels, thus making them very useful in hammer forging. In addition, they exhibit a resistance to heat checking and thermal softening equivalent to, or better than, chromium hot-work die steels tempered to the same hardness. These attributes as well as their ease of machining have led to their increasing use in forging of aluminum alloys, stainless steels, alloys steels, and nickel-base alloys. It appears that their only drawback is their somewhat low wear resistance, which can be more than compensated for by surface treatments such as nitriding (Ref 10, 11).

Cast Steels. The discussion up to now has dealt with properties and selection of wrought die steels. It appears that use of cast forging dies and die inserts is receiving renewed attention. This is surprising in view of early bad experiences with these types of die materials. Much of this revolution has been brought about by the use of ceramic or refractory casting mold materials, which can be preheated prior to pouring of the casting (Ref 12). With these types of mold materials, castings cool very slowly, and there is no surface chilling or development of columnar grain structures, the usual source of weakness in cast structures. Furthermore, gas bubbles do not become entrapped and there is much better chemical homogeneity than in the conventionally cast metals.

Cast tool steels have been shown to possess superior hot hardness and wear resistance when compared to their wrought tool steel counterparts, and they have been used extensively in presses as well as in hammers (Ref 13). Among the grades that have been used in cast form are H12, H13, H19, H21, H26, and H42 (Ref 14). It has also been shown that cast dies show improved resistance to heat checking and are tougher inasmuch as they are not as likely to crack in sharp corners (Ref 15). At present, because of the large initial cost of the pattern, it appears that cast dies are not economical unless at least three castings are needed from one mold. However, plant experiences showing increases in tool life in excess of 50% over wrought dies may justify their consideration in some instances.

Superalloys. Conventional forging of titanium and nickel-base alloys and isothermal, hot forging of these and other alloys place stringent requirements on dies and die materials. For this reason, die materials other than the conventional hot-work steels must be employed. Die materials such as the superalloys based on nickel, iron, and cobalt, which possess substantially greater high-temperature strength and resistance to thermal softening, have been found to be useful for these applications in which higher die temperatures are common.

Nickel-base superalloys used for forging dies come in cast and wrought forms. The cast superalloys generally have lower ductility and toughness compared to the wrought alloys but have been found to be useful in operations such as isothermal forging of titanium alloys where these attributes are not important. Among the cast superalloys most often employed are IN-100, Udimet 500, Nimocast 80, and 90, and Inconel 713LC. These alloys have hot yield strengths comparable to that of H11 at temperatures up to 540 to 650 °C (1000 to 1200 °F) (Fig. 4). In contrast to the hot-work die steels, they maintain large amounts of strength up to temperatures of 760 to 925 °C (1400 to 1700 °F).

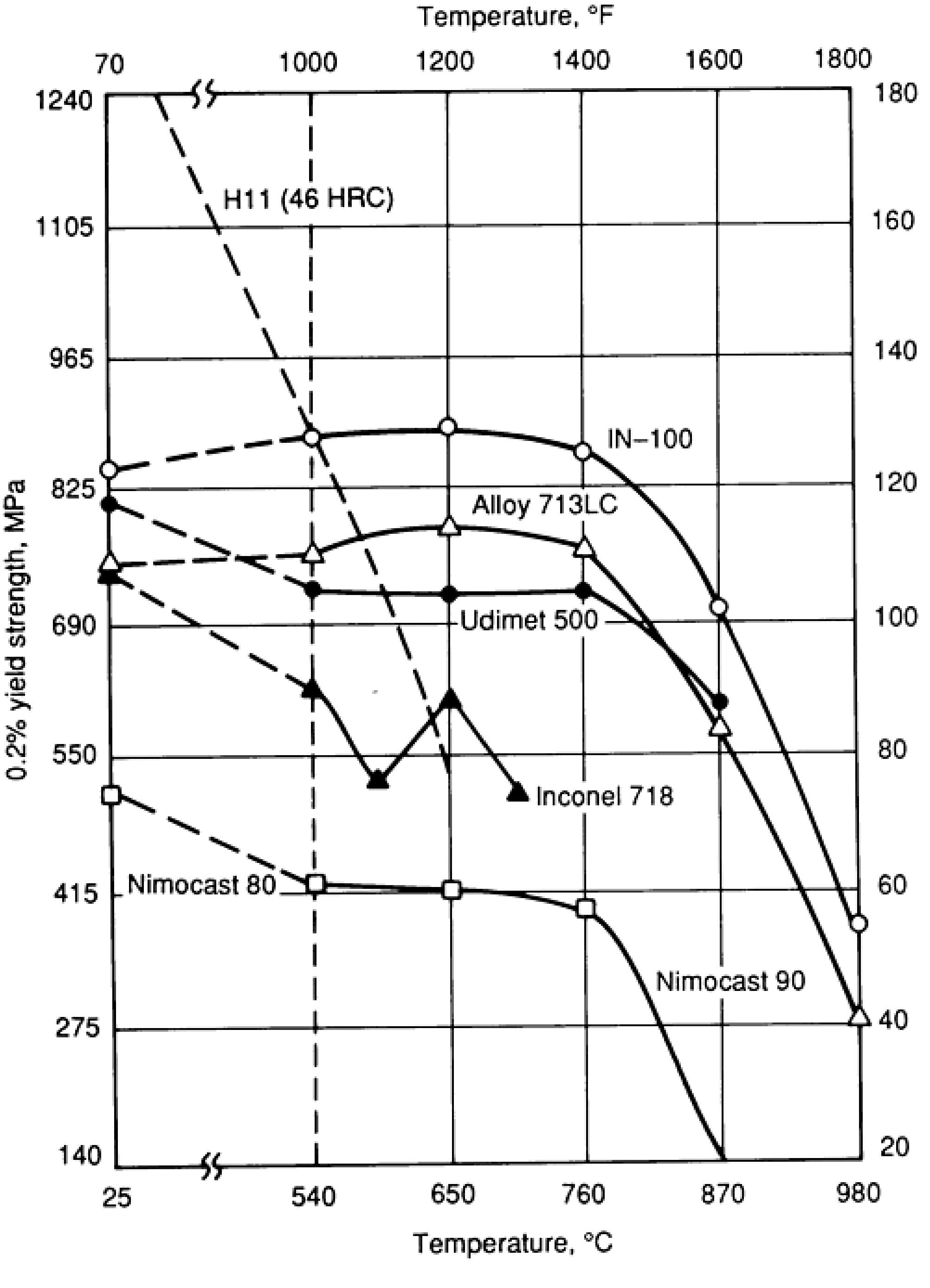


Fig. 4 Resistance of selected cast nickel-base superalloys to plastic deformation at elevated temperatures. H11 is included for comparison. Source: Ref 16

Wrought nickel-base superalloys usually have better overall mechanical properties, which make them favorable over their cast counterparts as far as selection of a forging die material is concerned. Typical wrought nickel-base forging die materials are Inconel 718, Waspaloy, and Udimet 700. Common iron-nickel alloys are Inconel 706 and Incoloy 901. As shown in Fig. 5, these alloys retain high hot strength up to temperatures of 650 to 870 °C (1200 to 1600 °F). As for the cast nickel-base superalloys, the drop in strength above these temperatures is related to γ' precipitate reversion. In the iron-nickel-base alloys, this drop in strength is due to reversion of other intermetallics, such as those based on niobium. The wrought superalloys have far better ductility and impact strength than the cast ones. This difference is particularly noticeable at temperatures between 540 and 870 °C (1000 and 1600 °F), where the superalloys have their greatest use as forging die materials. The values of ductility and impact energy of these wrought alloys are comparable to those of the hot-work die steels at the temperatures at which they are usually employed, namely 370 to 595 °C (700 to 1100 °F).

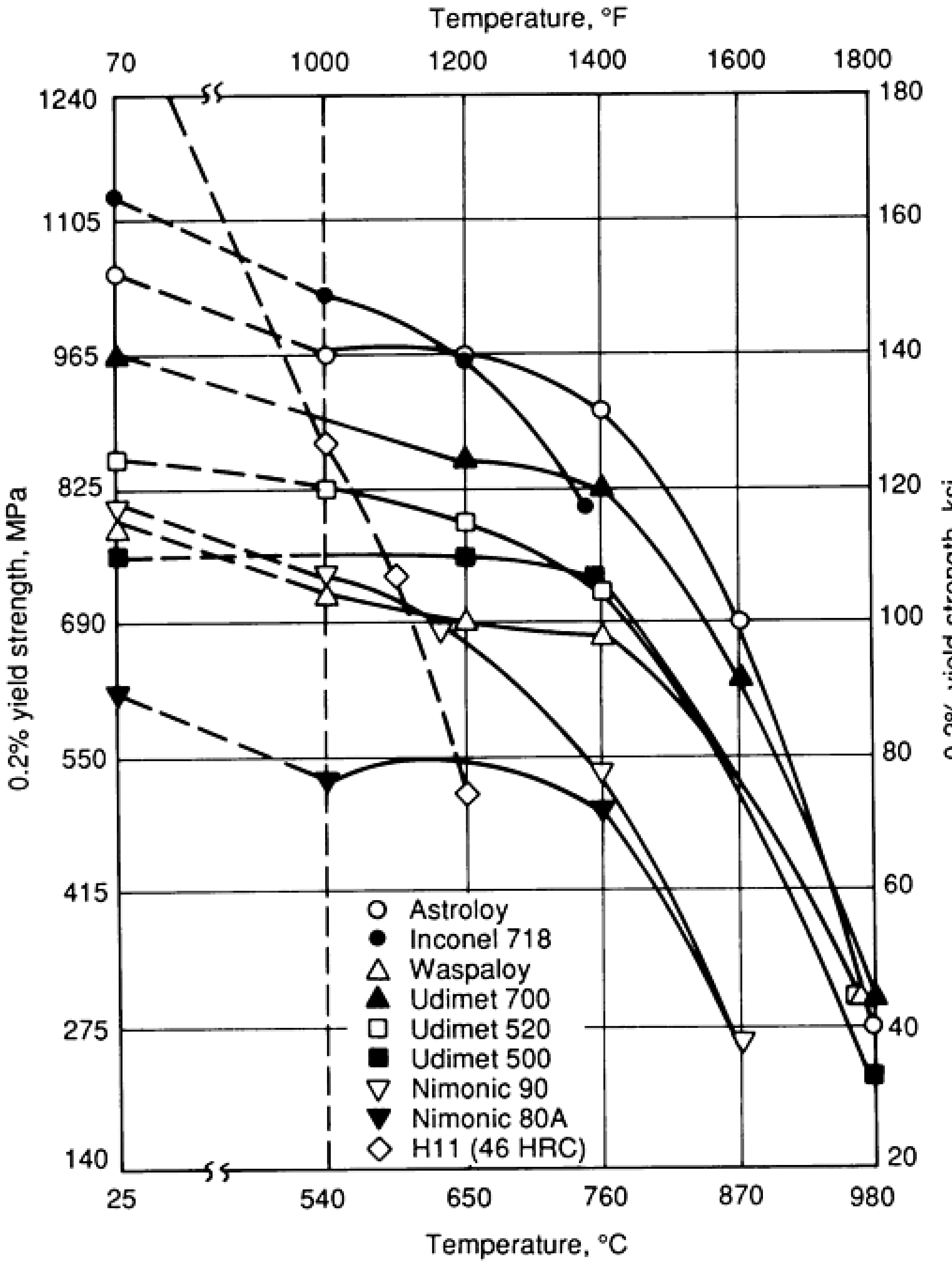


Fig. 5 Resistance of selected wrought nickel-base superalloys to plastic deformation at elevated temperatures. H11 is included for comparison. Source: Ref 16, 17

Cobalt-base superalloys, that find their main application in forging of titanium and nickel-base alloys, are less widely used for die materials compared with nickel- and iron-base superalloys. This is because of their higher cost. However, die wear studies on alloys such as Haynes Alloy 25 have shown that they outwear alloy steel dies by a factor of five to one in forging superalloys (Ref 18).

TZM Molybdenum. The forging of nickel-base superalloys places the greatest demands on forging die materials. Often these alloys must be forged isothermally at temperatures approaching and exceeding 1095 °C (2000 °F). In these cases, refractory metal alloys or ceramics must be employed as die material. Probably the most common alloys for these applications is *TZM* molybdenum (Ref 15). Because this alloy oxidizes very readily at temperatures around 1095 °C (2000 °F), tooling made of it must be enclosed in an evacuated chamber. Moreover, because of the high cost of the alloy and required peripheral equipment, it is used to forge only the most difficult-to-work alloys, which otherwise cannot be forged.

Cemented Carbides. Use of powder metallurgy materials such as tungsten carbide and titanium carbide as die materials, although fairly common in cold forging, appears to be rather limited in warm and hot forging. Carbides for warm and hot forging typically contain large amounts (10 to 20%) of cobalt binder to improve the toughness of the brittle carbide. Grade K3520, for example, is a tungsten carbide alloy that has 20% binder and that was designed specifically for hot working of steels (Ref 19). In addition to its excellent wear and shock loading resistance, it has thermal properties that minimize heat checking. However, because of the tendency of tungsten carbide to oxidize above 540 °C (1000 °F), its usefulness is limited. For die temperatures above 540 °C (1000 °F), titanium carbide materials appear to be more suitable. Although there is limited data on the use of titanium carbides, it appears that their major application is as die inserts in which large amounts of transverse compressive loading are applied to avoid the possible generation of tensile loads, which carbides generally cannot withstand without fracture.

Die Materials for Hot Extrusion

All extrusion tooling components that come in contact with hot billets have limited life and must be replaced periodically. Although the die is exposed to the most severe service, other die components such as backers, bolsters, mandrels, and dummy blocks may also deteriorate rapidly and are thus referred to as perishable tools. Containers, liners, liner inserts, and rams are exposed to less severe service and are known as durable tools. This arbitrary separation is used in Table 4, which summarizes recommendations for tool material and hardness. It will be noted that the 5% Cr grades of hot-work tool steel (H11, H12, and H13) predominate as tool material recommendations, especially for extruding aluminum and magnesium alloys. These tool steels are characterized by a high degree of toughness or resistance to breakage, and moderately high hot hardness to resist softening at elevated temperature. Steel H12 is the most widely used of these three, although H11 and H13 are successfully used. There are no data that conclusively prove the superiority of one of these steels over another, except when special techniques such as water cooling are employed. Although water-cooled tools are not used extensively, some plants have found it advantageous to water cool dies and mandrels, either internally or externally, especially when extruding aluminum and magnesium alloys. This practice is also used in copper, brass, and steel extrusion. For water-cooled tools, either H11 or H13 is recommended in preference to H12, because H12 is more likely to crack.

Table 4 Steels and hardnesses recommended for hot extrusion tools

Tool	Alloys to be extruded		
	Steel, titanium alloys, nickel alloys	**Copper and copper alloys**	**Aluminum and magnesium alloys**
Perishable tools			
Die	H11, H13, H21 at 43-47 HRC[(a)]	H21, H23, H26 at 36-45 HRC[(a)]	H12, H13, at 46-50 HRC[(a)]
Backer	H12 at 42-46 HRC	H12, H13[(b)] at 45-48 HRC	H12, H13[(b)] at 48-52 HRC
Bolster	H12 at 42-46 HRC	H12, H13[(b)] at 42-46 HRC	H12, H13[(b)] at 48-52 HRC
Mandrel	H11, H21 at 40-44 HRC	H11 at 40-46 HRC	H12, H13 at 48-52 HRC
Dummy block	H14, H21 at 40-44 HRC	H14, H21 at 40-44 HRC	H12, H13 at 44-48 HRC
Durable tools			
Container	4350 mod[(c)] at 300-350 HB	4350 mod[(c)] at 300-350 HB	4350 mod[(c)] at 300-350 HB
Liner	H12 at 400-450 HB, H21 at 375-400 HB	H12 at 400-450 HB, H21 at 375-400 HB	H12 at 400-450 HB
Liner insert	H12 at 400-450 HB, H21 at 375-400 HB	H12 at 400-450 HB, H21 at 375-400 HB	H12 at 400-450 HB
Ram	H12 at 450-500 HB	H12 at 450-500 HB	H12 at 450-500 HB

Note: Where more than one tool material is recommended for a specific purpose, listing is in order of increasing cost, the lowest-cost steel being shown first.

Source: Ref 3

(a) For dies of complicated shape, hardness is usually 4 or 5 points Rockwell C less than the values shown.

(b) Alloy steels such as 6150, 4150, and 4350 are occasionally used.

(c) Higher molybdenum content (usually 0.40 to 0.50%) than standard.

As indicated in Table 4, alloy steels such as 6150, 4150, 4350, and modified 4350 (with higher-than-normal molybdenum content) are sometimes used for tools such as backers and bolsters. Steel 4350 (unmodified or modified) is widely used for containers. These alloy steels are much cheaper than the hot-work tool steels (the base price of 4350 is only about one-fourth that of H12), but the difference in steel cost is hardly enough to show a significant savings in the total cost of

smaller backers or bolsters. For larger tools that have heavier sections, the alloy constructional steels are precluded for backers and bolsters because of inadequate hardenability. Except for containers, the use of low-alloy constructional steels is limited for hot extrusion tools.

The 5% Cr hot-work tool steels are also used in some instances for extruding metals that require higher temperatures than aluminum and magnesium. However, when the temperature of the metal being extruded is above 595 °C (1100 °F), the higher-alloy hot-work steels such as H21 are more widely used and recommended, especially for dies, mandrels, and dummy blocks.

Die Materials for Cold Heading

A large percentage of cold-heading dies can be made from a shallow-hardening steel such as the water-hardening carbon tool steels with or without chromium and vanadium. Steels W1, W2, and W5, with carbon contents from 0.85 to 1.10%, are usually used. Heading dies are also made from tungsten carbides with a cobalt binder. A WC-25Co material is widely used for heading work and has the distinct advantage of being machinable: it can be bored, drilled, and turned with carbide tools. Dies that require greater wear resistance and that are subjected to less shock are made with 13 to 16% Co binder. These grades must be ground or lapped with diamond. Carbide dies are usually of the insert type, with a hardened steel case supporting the carbide insert. Although solid carbide dies have been used, unsupported dies of this type are extremely rare.

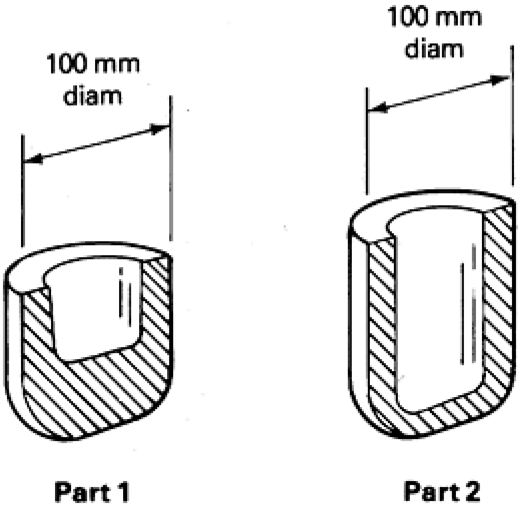
Die Materials for Cold Extrusion

Compressive strength of the punch and tensile strength of the die are among the most important factors influencing the selection of material for cold extrusion tools. Because the die is invariably prestressed in compression by the pressure of inner and outer shrink rings, the principal requirement for a satisfactory die is a combination of tensile yield strength and prestressing that will prevent failure. Punches require sufficient compressive strength to resist upsetting without being hazardously brittle. Thus, almost without exception, and particularly for extruding steel, the primary tools in contact with the work must be made from steels that will through harden in the section sizes involved. This is notably different from cold-heading tools, in which a hard case and soft core are usually desired.

Table 5 shows some typical recommendations for the punch, die, and knockout for two simple backward-extrusion operations. Both operations are similar in severity when part 1 is made from a cylindrical slug and part 2 is made from part 1 (in aluminum, part 2 can be made in a single step directly from a cylindrical blank).

Table 5 Recommended tool steels for backward extrusion of two parts

Metal to be extruded	Total quantity of parts to be extruded(a) 5,000	50,000
Punch material		
Aluminum alloys	A2	A2, D2
Carbon steel, up to 0.40% C	A2	D2, M2(b)
Carburizing grades of alloy steel	A2	M2(b)
Die material		
Aluminum alloys	W1(c)	W1(c)
Carbon steel, up to 0.40% C	O1, A2	A2(d)
Carburizing grades of alloy steel	O1, A2	A2(d)
Knockout material		
Aluminum alloys	A2	D2
Carbon steel, up to 0.40% C, and carburizing grades of alloy steel	A2	A2, D2



Note: Where two tool materials are recommended for the same conditions they are given in order of cost—the less expensive being shown first. (a) For part 1, starting with a solid slug; for part 2 starting with part 1. In aluminum, part 2 can be made directly from a cylindrical blank. (b) Liquid nitrided. (c) The 1.00% C grade is recommended. (d) Gas nitrided on the inside diameter only. Source: Ref 3

Source: Ref 3

The recommendations for dies to extrude these parts from steel are conservative, because a D2 punch might achieve a total life of 300,000 pieces, with 60,000 between redressings; for the dies, O1 might achieve 40,000 pieces between redressings and 160,000 total, as compared with 70,000 and 200,000 pieces, respectively, for A2 steel.

Because of economic considerations, sintered carbide should be considered for punches for long runs (over 500,000 parts) wherever press and die equipment is rigid enough not to cause breakage of carbide tools. When the wear limit of tools (or the part tolerance) is narrow, or when runs are long and production shutdowns must be avoided, carbide is used as insert material in the dies to extrude either steel or aluminum.

Die Materials for Sheet Metal Forming

The useful performance of a forming die is measured in terms of its wear. Total wear is affected primarily by the length of the production run and the severity of the forming operation. Hence, production quantity and severity are the most important factors in selecting die materials for forming. The metal being formed, its thickness, and the finish and dimensional tolerances required in the part are also influential factors.

Recommended die materials range from plastics for low reduction of simple to moderate parts, up to the most wear-resistant tool steels, surface hardened by nitriding, for making severely formed parts (Ref 3). Parts of even greater severity, or those run in quantities larger than one million, may require dies or inserts of cemented carbide.

Reference 3 gives a number of typical tool steels and performance data. Cast iron is a useful die material for forming parts larger than about 300 mm (12 in.). Its performance makes it suitable for use in medium production runs or in short runs of large parts. When cast iron is used with inserts, it will produce greater quantities. Cast iron should have predominantly fine pearlite with no massive carbides and a minimum of ferrite. Graphite should be of AFS-ASTM type A distribution with a preferred flake size of 4 to 5.

The zinc alloy recommended for some sheet forming dies has high pattern fidelity when cast, but requires very accurate shrink patterns to minimize hand labor in finishing. It is economical only for die components at least 300 mm (12 in.) long, and is most economical in dies about 900 mm (36 in.) long. One of its principal values is for complicated dies that would require intricate machining and hand tooling if made of cast iron, but which can be cast more closely with zinc. In tensile strength, compressive strength, and hardness, zinc alloy tools are inferior to other metals used in die construction. Nevertheless, a zinc alloy containing 4% Al, 3.25% Cu, 0.1% Mg, 0.7% Ni, and 0.2% Ti has provided die life up to three times that of the standard zinc alloys.

Where a limited variety of tool materials are available, hot-rolled mild steel plate with carbon content from 0.10 to 0.20% is in major use as a sheet-forming die material, and is also widely used elsewhere for shot-run forming of small parts. This type of steel is not recommended without surface hardening except for quantities of less than 10,000 parts. In addition, hot-rolled mild steel can be used for dies only where straightening facilities are available for correcting distortion induced by heat treatment.

Several medium-carbon alloy steels such as 4140 are available in plate form and are useful for some types of forming dies. For example, steels 8640 and 5140 are suitable alternates for 4140 and cost about the same.

Cast steel is used for forming parts larger than 300 mm (12 in.) long where the shape of the part makes a casting more feasible than wrought stock. Plain carbon steel (0.65 to 0.85% C) and alloy cast steel (0.40 to 0.50% C, 0.90 to 1.25% Mn, 0.90 to 1.25% Cr, and 0.35 to 0.50% Mo) are two commonly used compositions. Cast steel is more costly and more difficult to machine than cast iron and is more likely to gall. However, it is tougher than cast iron. One important advantage of cast steel, particularly the plain carbon grade, is its weldability. It can be rewelded with steel or hardfaced with tool steel, aluminum bronze, or other hardfacing material. Because of its poorer resistance to galling, it is less desirable for dies for forming carbon and stainless steels than for those for forming aluminum and copper alloys. Cast steel is also useful for restrike, flanging, and other types of dies that are less likely to gall or pick up material from the sheet.

Aluminum bronzes, which are alloys principally of copper and aluminum, are available in hardnesses ranging from 120 to 340 HB. These alloys have excellent resistance to galling and are desirable for dies where best finish is required on carbon and stainless steel parts.

For long runs, inserts of sintered carbide are now widely used, especially for deep-drawing dies. In dies of about 200 mm (8 in.) or less for continuous production of over 1,000,000 pieces, carbide has in many instances proved to be the most economical die material. Such dies have maintained size in drawing 60% reductions of more than 500,000 pieces and have made as many as 1,000,000 parts with reductions greater than 40% when the steel to be drawn was surface treated with zinc phosphate and soap. However, carbide is not superior to tool steel such as D2 in complex deep-drawing operations, such as those with reductions greater than 40% that combine drawing with coining or stretching.

Even more than with other tooling materials, the quality and economy of plastic tooling improves with shop experience and volume. Polyester, epoxy, phenolic resin, and nylon have been used. Plastic dies that achieve and longest life are constructed so that the wearing surface is faced with glass cloth that has had most of the plastic material forced out under

pressure before and during curing. Plastics are most economical when a model is available as a pattern for the lay-up of the plastic-impregnated glass cloth facing, which is backed up with chopped glass fibers impregnated with 50% resin. The strength (70 to 275 MPa, or 10 to 40 ksi) and, in particular, the hardness of plastics are inferior to those of metallic die materials. In some instances, the problem of obtaining the required strength in a plastic die must be solved by extra work in design or development. In many dies, however, the regions of high pressure are confined to local areas, which, if anticipated, are strengthened with metallic inserts.

Materials for Die-Casting Dies

The important properties required of materials for die-casting dies are resistance to thermal shock and to softening at elevated temperatures. Resistance to softening is required to withstand the erosive action of molten metal under high injection velocity. Other properties that influence selection of materials for die-casting dies are hardening characteristics, machinability, resistance to heat checking, and weldability. Availability and cost are also important, especially for dies larger than about 200 mm (8 in.).

The performance of die-casting dies is directly related to the casting temperature of the work metal, thermal gradients within the dies, and the frequency of exposure to high temperature. Tool steels of increasingly higher alloy content are required as the casting temperature increases. As indicated in Table 6, hot-work tool steels are not recommended for casting zinc except for long production runs.

Table 6 Recommended materials for die-casting and die inserts

Allow to be cast	Number of shots		
	50,000 or less	**250,000**	**1,000,000**
Zinc (25 mm die cavity)	P20[a][b]	P20[a][b]	P20[b], H13[b]
Zinc (100 mm die cavity)	P20[a][b]	P20[a][c], 4150 mod[d]	4150 mod[d], H13[b]
Aluminum or magnesium	H11, H13	H13, H11	H13, H11
Copper	H21, H20, H22, H23	. . .	. . .

Source: Ref 3

(a) Prehardened to 280 to 320 HB; SAE 4140, prehardened, can be used when best surface finish is not required.

(b) Recommended for die inserts.

(c) Hardenability barely adequate for these conditions.

(d) Tool steel quality modified to desired hardenability in prehardened condition.

Die hardness for zinc casting is less exacting than for alloys of higher lasting temperature, which usually permits the use of steels prehardened by the manufacturer to a maximum hardness consistent with suitable machinability. A typical hardness range is 29 to 34 HRC. Also under development and in preliminary commercial use are precipitation-hardening steels for die-casting dies. In addition to being hardenable, they have superior machinability at higher hardness than quench-hardening tool steels.

Hot-work tool steels are almost always used for casting the higher melting point alloys, such as aluminum and copper. For aluminum and magnesium alloys, the H11 and H13 steels are hardened to about 44 to 48 HRC; the H20, H21, and H22 steels for casting copper alloys are generally used in the hardness range from 38 to 45 HRC.

Forging steels of tool steel quality and increased hardenability occasionally are used for applications requiring die blocks so large as to be unavailable in hot-work tool steel. However, the use of these high-hardenability steels is limited by the availability of the forgings and greater difficulty in welding them. Sometimes unhardened low-alloy steels are used for dies of simple design and for short runs of noncritical parts.

Wear and Failure in Die-Casting Dies

The die-casting process uses internally cooled metal dies to mold aluminum, zinc, brass, and other materials into intricate cast shapes. It provides a rapid, close-tolerance, and high-production method for producing large volumes of parts at a relatively low cost--for example, automotive parts such as carburetor bodies, transmission cases, and valve bodies.

Die life is a major consideration in a die-casting process, because, depending on the complexity of the part being produced, a die may cost more than the die-casting machine itself. Die lives may vary from 20,000 to over 250,000 parts depending on die design, die material, workpiece material, and the thermomechanical processing involved.

A typical design for a die-casting die is shown in Fig. 6. It includes two sections: the fixed die half and the ejector die half. The fixed die half is mounted toward the liquid metal injection system. The ejector die half is mounted on a movable platen of the machine. The casting, on cooling, sticks to the movable die half and is ejected via ejector pins mounted on an ejector plate. Guide pins ensure proper alignment of the two halves. Die inserts contain the outside geometry of the part in a cavity into which the molten metal is injected under pressure. Fixed or removable cores are positioned at positions where holes in the casting are desired. Heat is removed from the die by circulating cold water through cooling vents in the dies.

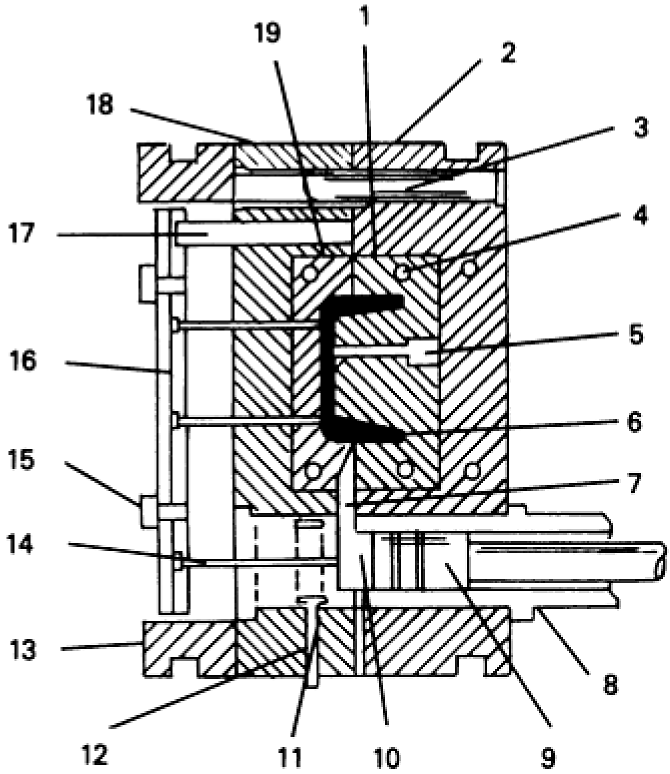


Fig. 6 A typical die-casting die. (1) Fixed-half tool insert, (2) fixed-half tool holder, (3) guide pin and bush, (4) water cooling, (5) fixed core pin, (6) casting, (7) runner, (8) shot sleeve, (9) plunger tip (in forward position), (10) slug, (11) slug cooling, (12) water inlet, (13) ejector box, (14) ejector pins, (15) ejector plate stop, (16) ejector plates, (17) ejector plate return pin, (18) moving-half tool, (19) moving-half tool insert. Source: Ref 20

During its service life, the die cavity (inserts) and cores experience a large number of thermomechanical cycles. A network of fine cracks develops on the die surface due to thermal fatigue. This failure mode is called heat checking and can cause rejection of the casting. The die may be repaired by grinding or welding, or discarded if gross cracking occurs due to thermal shock. In addition to heat checking and gross failure, other failure mechanisms include abrasive wear, adhesive wear, stress-corrosion, and erosion due to liquid metal attack. With the exception of molten zinc, investigators have found very little evidence of reaction between molten materials and die steels. Table 7 shows some of the factors that limit die lives for different casting metals.

Table 7 Life-limiting factors for die-casting dies

Casting alloy	Factors that limit die life	Normal life, number of shots × 10^3	
		Die	**Core**
Zinc	Erosion, mechanical abrasion	500-1000	500-1000
Aluminum	Heat checking, cracking, erosion, low compressive strength	100-250	50-150
Copper (brass)	Heat checking, low compressive strength, erosion, cracking	10-100	3-20

Wear and Failure Modes in Die Casting

Wear and failure of die-casting dies involve a complex interaction between various failure modes. Some of the important wear and failure modes are (Ref 21, 22, 23, 24, 25, 26, 27, 28, 29):

- Erosive and abrasive wear
- Chemical attack, or corrosion
- Thermal fatigue, or heat checking
- Soldering
- Gross cracking, including thermal shock

Thermal fatigue is the most important failure mode, followed closely by soldering. These two modes often cause frequent shutdowns and redressing of the dies. Erosive wear, which depends on the impinging velocity and the angle of impact, can be minimized substantially by proper gating design (Ref 23). Chemical attack (corrosion) is important in that it influences the initiation and growth of heat-checking cracks. Gross cracking, however, is primarily due to mechanical or thermal overload (thermal shock) and can be mitigated by cavity redesign. These failure modes will be further described below.

The moving parts of the die-casting die (cores, slides, and pins) are required to have all the properties of the stationary parts, and they must also be resistant to wear. The wear of die-casting dies can be reduced by (Ref 22):

- Nitriding one or both of the surfaces of contacting steels
- Using lubricants
- Shallow notching or dimpling of the slide or gib
- Maintaining a proper fit between wearing couples
- Providing polished surfaces on mating members

Erosive Wear. Wear in die-casting dies is primarily due to erosion of the die surface, which involves the washing away of die material by the impinging jet of an incoming molten metal stream. Erosion is a function of gating and casting techniques and properties of the die material. The impingement velocity and angle of molten metal jet against a core or die face at the gate inlet can be reduced by proper die design (Ref 30).

For casting copper alloys, the moving parts and the cavity that come in contact with molten metal are made from the same hot-working die-steel. Parts not in contact with the melt can be made of alloy steel such as 6145, which has good hot hardness.

Test for Erosion and Corrosion Evaluation. In die casting, it is often difficult to differentiate between erosion and corrosion phenomena, and the die damage mechanism may actually be a combination of erosion and corrosion. A laboratory test method has been developed at the Uddeholm Steel Research Department in Sweden to evaluate erosion-corrosion resistance of die steels (Ref 21). A schematic of the equipment setup is shown in Fig. 7.

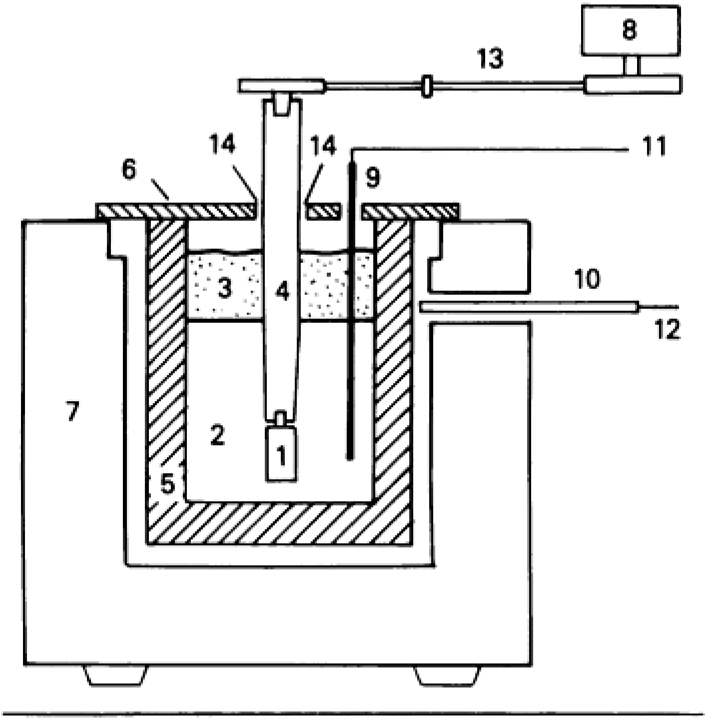


Fig. 7 Equipment for erosion-corrosion test in various molten metals. (1) Testpiece, (2) molten zinc, (3) graphite, (4) graphite bar, (5) graphite crucible, (6) refractory cover, (7) furnace, (8) motor, (9) and (10) thermocouples, (11) to potentiometer, (12) temperature regulator, (13) chain drive, (14) inlet orifice for nitrogen gas. Source: Ref 21

The erosive attack of the melt on the die surface is influenced by both the material properties and process parameters, including steel grade, heat treatment, surface treatment, melt composition, and temperature. Weight loss due to erosion-corrosion in a large number of die steel samples was measured for zinc, aluminum, and brass melts. In a commercial zinc melt at 500 °C (930 °F), the losses were relatively small, but at 600 °C (1110 °F), most steel grades reported significant weight loss differences as shown in Fig. 8. In an aluminum melt, wear losses for soft annealed dies were significantly higher than those for hardened dies (Fig. 9). Brass melts were found to react more strongly in steels with higher nickel content (Fig. 10).

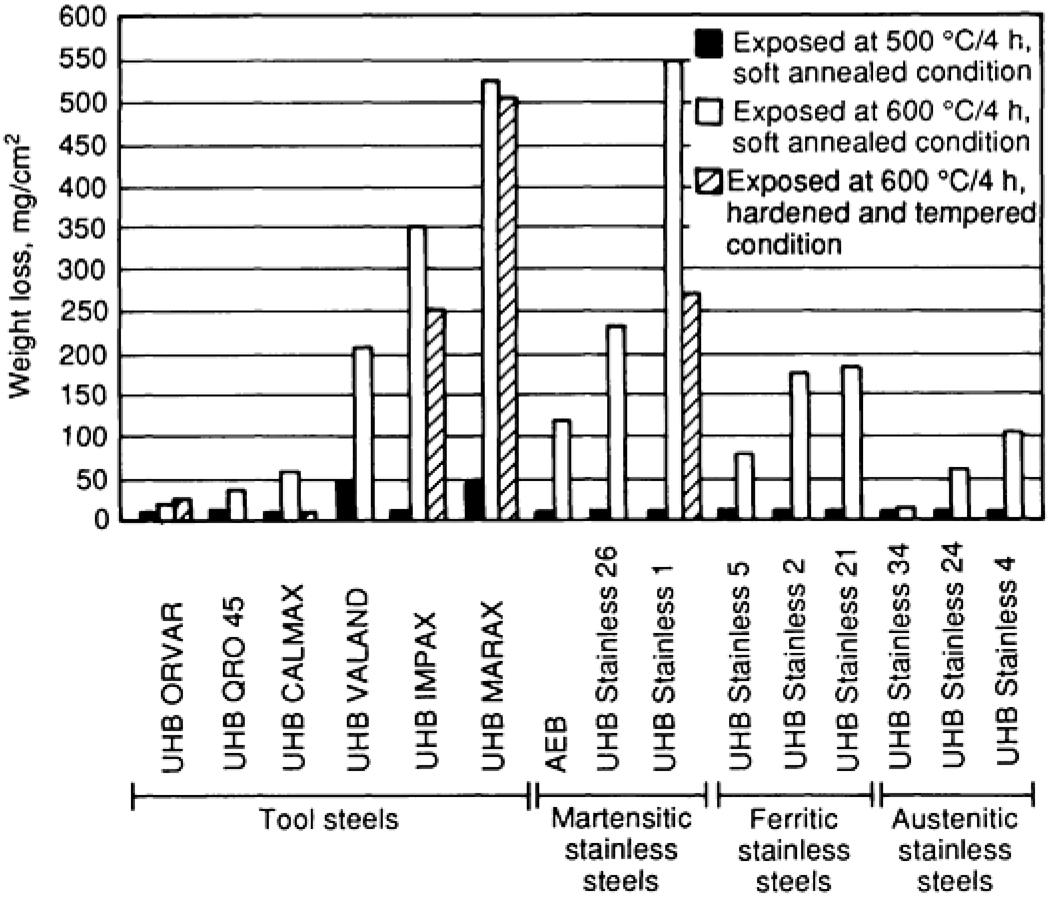


Fig. 8 Weight-loss measurements for different Uddeholm Steel die steel grades tested in a zinc (Zn-4Al) melt. Source: Ref 21

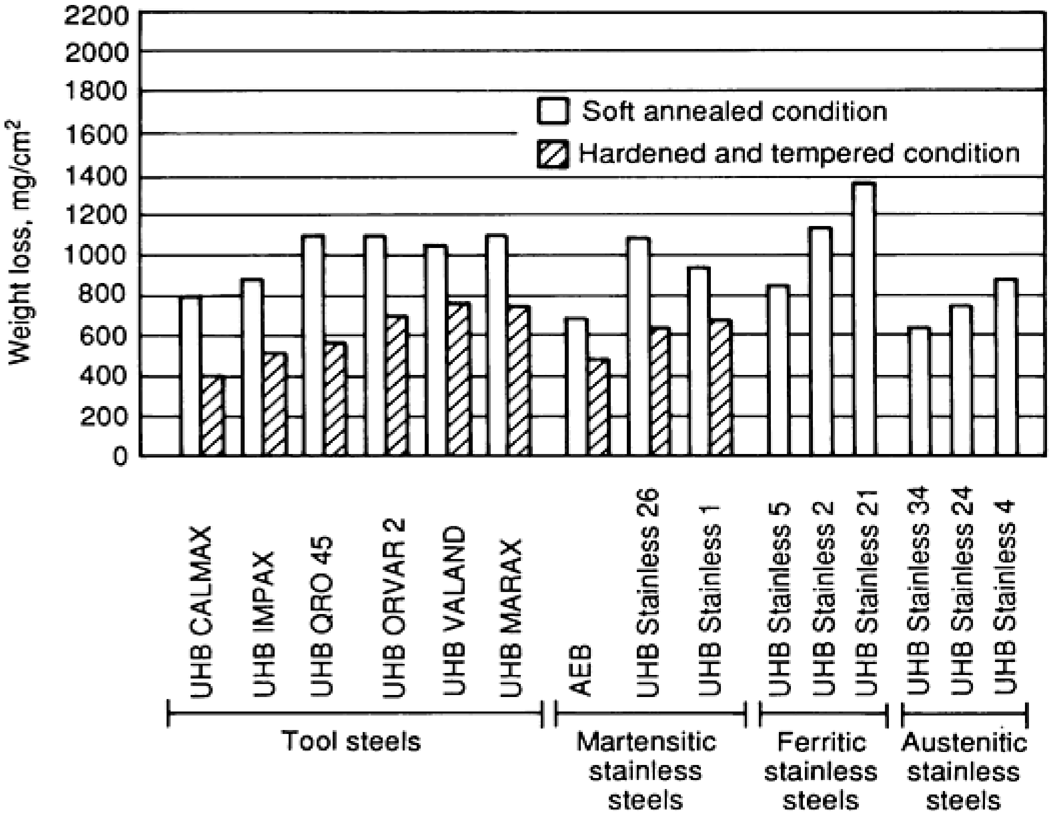


Fig. 9 Weight-loss measurements for different Uddeholm steel grades tested at 735 °C/4 h (1355 °F/4 h) in an aluminum (Al-9.9Si-0.8Fe) melt. Source: Ref 21

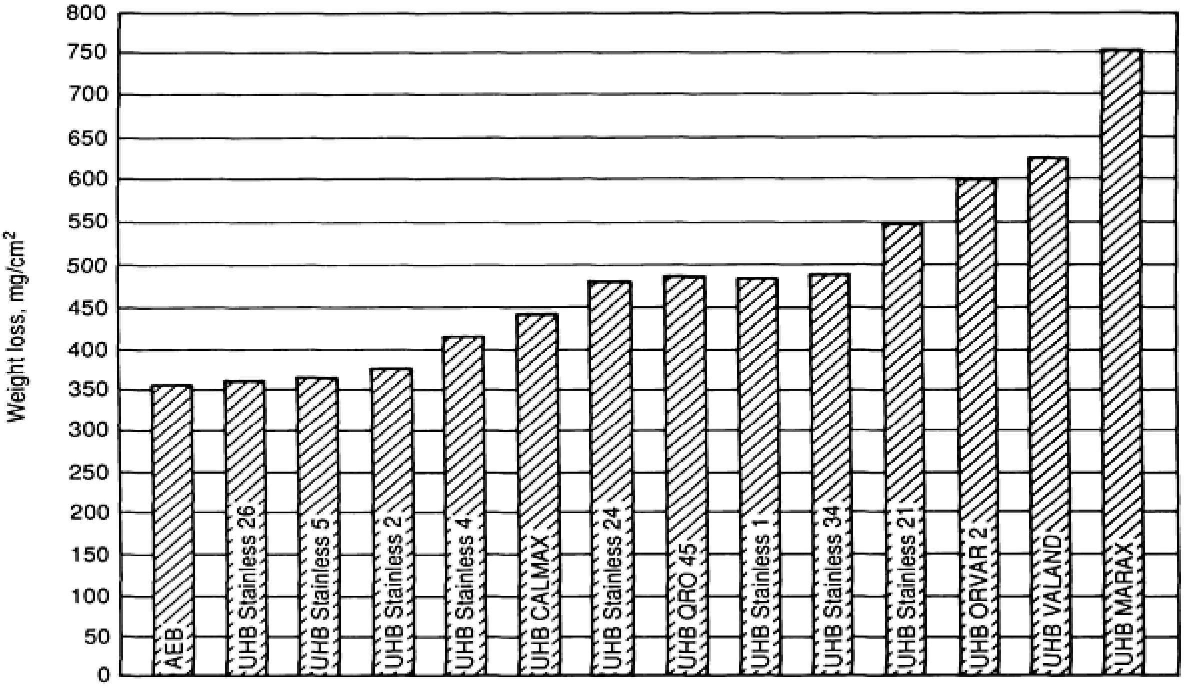


Fig. 10 Weight-loss measurement for different Uddeholm steel grades tested at 950 °C (1740 °F) for 4 h in a brass (Cu-38Zn-1Pb) melt. Source: Ref 21

Soft-annealed and hardened-and-tempered steels were also tested (Fig. 11). Hardened-and-tempered steels showed lower weight losses in zinc and aluminum melts. The higher alloy content in hardened-and-tempered steels reduces molten metal attack.

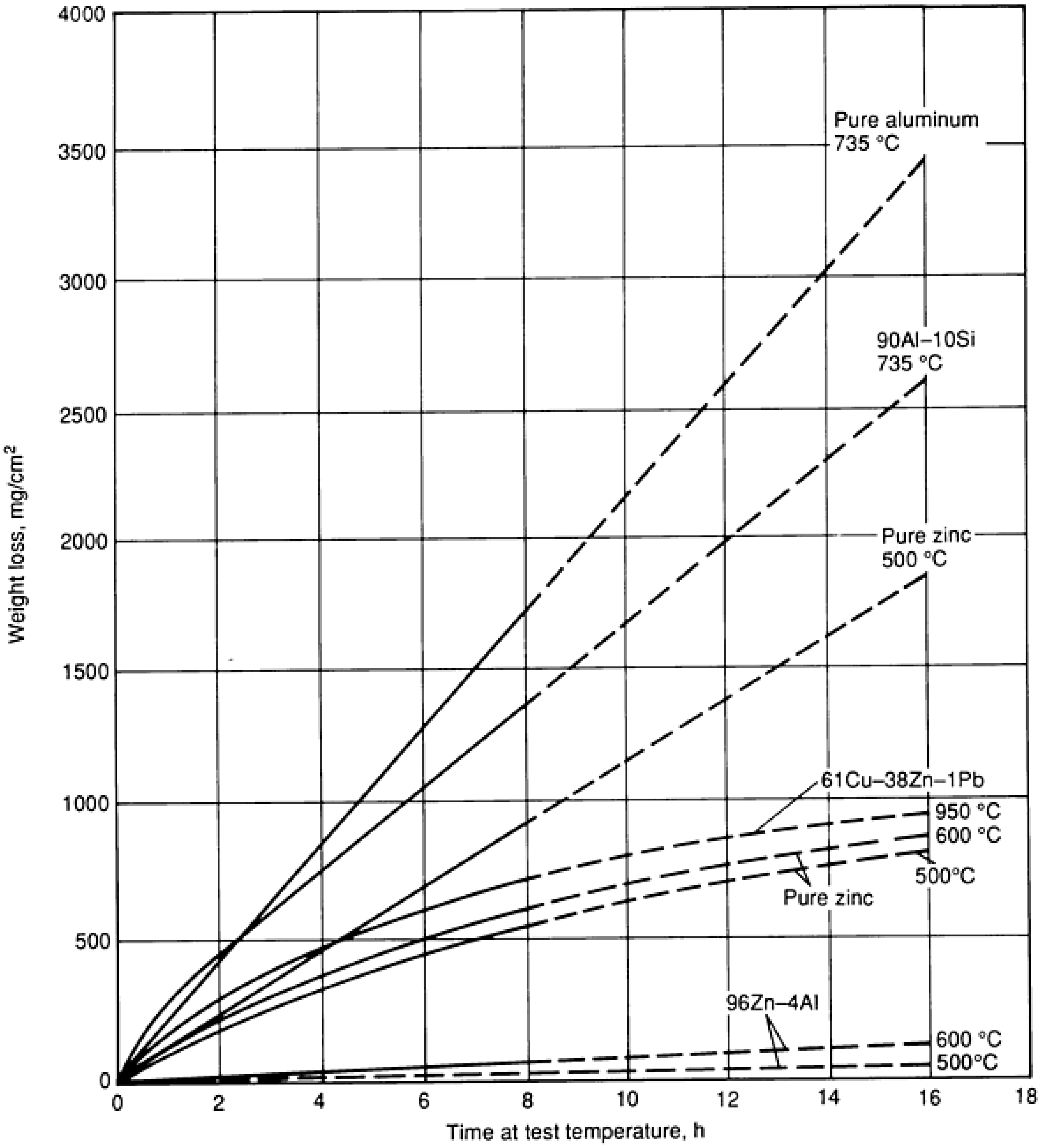


Fig. 11 Erosion rates of soft-annealed Uddeholm UHB Orvar 2 (AISI H13) samples tested in different melts. Source: Ref 21

A large number of different surface coatings were tested. It was found that erosive wear is substantially reduced if the protective oxide layer does not detach from the die surface during the casting process. Similar conclusions are reached with nitrided surfaces.

Melt composition also influences erosion-corrosion. Figure 11 shows large differences in weight loss caused by different melts. In zinc and aluminum melts, wear loss increases rapidly; in the case of brass melts, the wear loss increases gradually with melt temperature and time.

The following general conclusions can be drawn regarding erosion (Ref 21):

- Pure zinc melts are more prone to problems than commercial zinc alloys. The melt composition also seems to be a very important parameter for aluminum and brass melts
- Erosion attack generally decreases with decreasing melt temperature. Melt temperatures should be kept as low as possible, for example, zinc alloys ≤480 °C (≤895 °F), aluminum alloys ≤735 °C (≤1355 °F), and copper alloys ≤1000 °C (≤1830 °F)
- There appears to be no significant difference in erosion loss among different common hot-work tool steels
- Hardened-and-tempered samples generally show smaller losses than soft-annealed samples
- If metallic contact between the tool steel and the melt (such as by an oxide film) can be avoided, the risk of severe erosion attack decreases significantly

Heat checking in die casting is often caused by thermal fatigue, surface stresses, low material strength, and surface irregularities. During the process cycle, alternate heating and cooling leads to thermal fatigue. Mechanical and thermal stress fluctuations initiate fine cracks on the cavity surface that grow larger and ultimately lead to failure of the die. A considerable body of research on the resistance of die materials to heat checking and thermal fatigue has also been generated at Uddeholm Steel. Most noteworthy are a qualitative heat checking rating diagram scale (Ref 21, 30) and a laboratory test for determining thermal fatigue (heat checking) properties of hot-work tool steels (Ref 21, 31, 32).

Materials Resistant to Heat Checking. To resist heat checking, die materials should have a low coefficient of thermal expansion, high thermal conductivity, high hot yield strength, good temper resistance, high creep strength, and adequate ductility (Ref 31). The coefficient of thermal expansion governs the magnitude of strains and stresses produced due to temperature changes and is strongly dependent on the type of microstructure of the die material. All steels with martensitic, bainitic, and ferritic microstructures have approximately the same thermal expansion coefficient. Steels with austenitic microstructure have 50% higher expansion coefficient values and are not suitable for die-casting applications. High thermal conductivity reduces thermal gradients by permitting rapid heat transfer from hot to cold locations. Lower temperature gradients produce lower thermal stresses. The hot yield strength governs the amount of plastic strain for a given temperature cycle. Heat checking, a low-cycle fatigue phenomenon, is controlled basically by plastic strain amplitude. Therefore, high hot yield strength is desirable. To maintain high hot yield strength at prolonged exposure to high temperature, good temper resistance is necessary. Temper softening, which is often accentuated by superposed mechanical loads, can be controlled by good creep strength. Because ductility is a measure of a material's ability to resist plastic strain without cracking, adequate ductility is essential for resistance to heat checking.

Thermal shock is gross cracking due to sudden temperature changes (Ref 30). It is a macroscopic phenomenon and often results in total failure of the dies. Thermal shock is primarily resisted by good material toughness and ductility and, to some extent, by high hot yield strength. A test method developed at Uddeholm Steel for evaluating thermal shock resistance is described in Ref 30 and 31.

Factors Influencing Wear and Failure in Die-Casting Dies

The factors influencing different wear and failure modes can be grouped as follows:

- Geometric factors in the design
- Die material factors, including heat treatment
- Interface factors, including lubrication, surface treatments, and coatings
- Processing factors, including heating and cooling cycles

The geometric factors in die design govern the mechanical stress and thermal gradients that develop in the die. Some important die design considerations are (Ref 31):

- Cavity should not be too large for the die block--there should be enough material left in the die to provide adequate strength and rigidity
- To minimize chilling, the die impression should not be too close to the water-cooling chamber
- Location of gating, venting, offsets, undercuts, deep narrow recesses, cores, and inserts should be optimized to reduce stress or thermal gradients
- Irregular cavity contours and sharp corners should be adequately filleted to reduce stress concentrations
- Location of large areas of molten metal in close proximity to smaller areas should be avoided

Part tolerances also influence die lives. Tighter part tolerances reduce die lives by permitting less wear on the dies. Surface finish requirements of the part can also limit die usage.

Die Material Considerations. In order to resist failure, die materials should have uniform microstructure, no internal defects, good machinability, good polishability, good response to heat treatment, good toughness, resistance to wear, and resistance to heat checking. Because H13 is the most commonly used die steel (H11 and H12 are sometimes used) for die casting, the effect of the material is limited to its manufacturing technique and the heat treatment. Although air-melted steels are less expensive, vacuum arc remelted (VAR) steels have better toughness and higher hot yield strengths (Ref 27). The austenitization and homogenization heat treatments have to be tightly controlled to achieve the most desirable microstructure. Rapid quenching is often prescribed to obtain martensitic microstructure. However, stepped cooling to lower bainite is often preferred because of higher toughness values (Ref 23).

Interface Considerations. Use of commercial die lubricants or mold release agents reduces soldering or sticking of the casting to the die. This subsequently reduces die wear. Some lubricants, however, reduce die life by chemically attacking die steels and also the casting metal adhering to the die surface. An effective lubricant permits easier cleaning of the dies, reduces wear, and allows longer runs between die polishes.

Surface treatments such as nitriding are often used to reduce erosive wear. A nitriding treatment can be substituted for the third, or final, temper of the dies (Ref 30). The surface hardness of H13 when gas nitrided can be 65 to 70 HRC. This extra hardness is desirable for die-casting dies that have been tempered to a lower hardness for increased toughness; the hard surface resists wear but not necessarily heat checking. The original hardness of the interior is not lowered because nitriding is done in the temperature range of 525 to 540 °C (975 to 1000 °F), which is usually below the tempering temperature of H13 steel. However, the nitrided case depth should not exceed 0.13 mm (0.005 in.); deeper cases can result in flaking or spalling damage at parting lines and sharp edges. Spalling of a nitrided surface (which causes undercuts in the spalled area) leads to tearing of the casting during ejection and to damaged cores. The use of a solid release agent lessens soldering, spalling, and abrasion. The use of molybdenum disulfide or colloidal graphite in oil also reduces wear.

Nitriding is especially beneficial in reducing abrasive wear due to sliding of moving components such as ejector pins, slides, and cores (Ref 33). However, the core and slide ends that contact molten metal must be masked against nitriding so as to obtain better resistance against heat checking and spalling. Ejector pins are available made from either nitrided H11 or nitriding steels such as 7140.

Sputtered coatings have also been evaluated for wear and heat checking resistance (Ref 32, 34, 35). Platinum, molybdenum, and tungsten have good resistance to thermal fatigue. They also inhibit oxidation of the substrates, which delays crack initiation. But once the coatings crack, heat checking of the exposed substrate is often higher than that of the virgin material due to the preferential site of metal attack. Carbide coatings such as niobium carbide or vanadium carbide have been successfully applied to die casting dies via the Toyota Diffusion (TD) process. These coatings have reduced die wear substantially in laboratory and selected production tests (see the section on "Surface Treatments and Coatings" in this article).

Processing conditions such as imposed mechanical and thermal loading histories influence die failure modes with more severe loadings resulting in increased wear and failure rates. Die preheating and machine locking power also affect die service lives (Ref 22).

Wear in Sheet Metal Forming Dies

The amount of wear of dies during sheet metal forming is proportional to the distance the sheet metal slides over the die for a given pressure between the surfaces in contact. Thin annealed sheet metal exerts the least pressure and thus causes the least wear; thick hardened sheet metal causes the most rapid wear. However, the rate of wear for different combinations of die/sheet metal pair may vary considerably, depending on material surface characteristics, the speed of forming, and the type of lubrication.

In the forming of parts with dies that produce wrinkles, high localized pressures develop on the tools. In the ironing stage, these wrinkles produce prohibitively high rates of wear and galling.

Sheet metal working consists of blanking or cutting, forming of shallow parts, and deep drawing. Wear of blanking dies will not be considered in this article but wear occurring during shallow forming and deep drawing will both be considered separately. The latter two operations impose different stress fields and wear profiles on the forming dies. Some recommendations are made concerning die material selection based on the expected die lives.

Information provided in this section is primarily from Ref 36 and 37. Reference 38 describes methods for testing mechanical behavior of steel sheet coated with corrosion-resistant coatings (zinc-aluminum); Ref 39 details a method for wear data processing. The topics described in Ref 38 and 39 will not be considered in this section because the focus is on die wear.

Wear in Shallow Forming Dies

To understand wear in shallow forming dies, it is essential to study the mechanics of the process based on the true shape of the parts to be drawn. For this purpose, six sample part shapes are included in Fig. 12, ranging from the simple shapes such as those of parts 1 and 2, to moderately severe shapes such as those of parts 5 and 6.

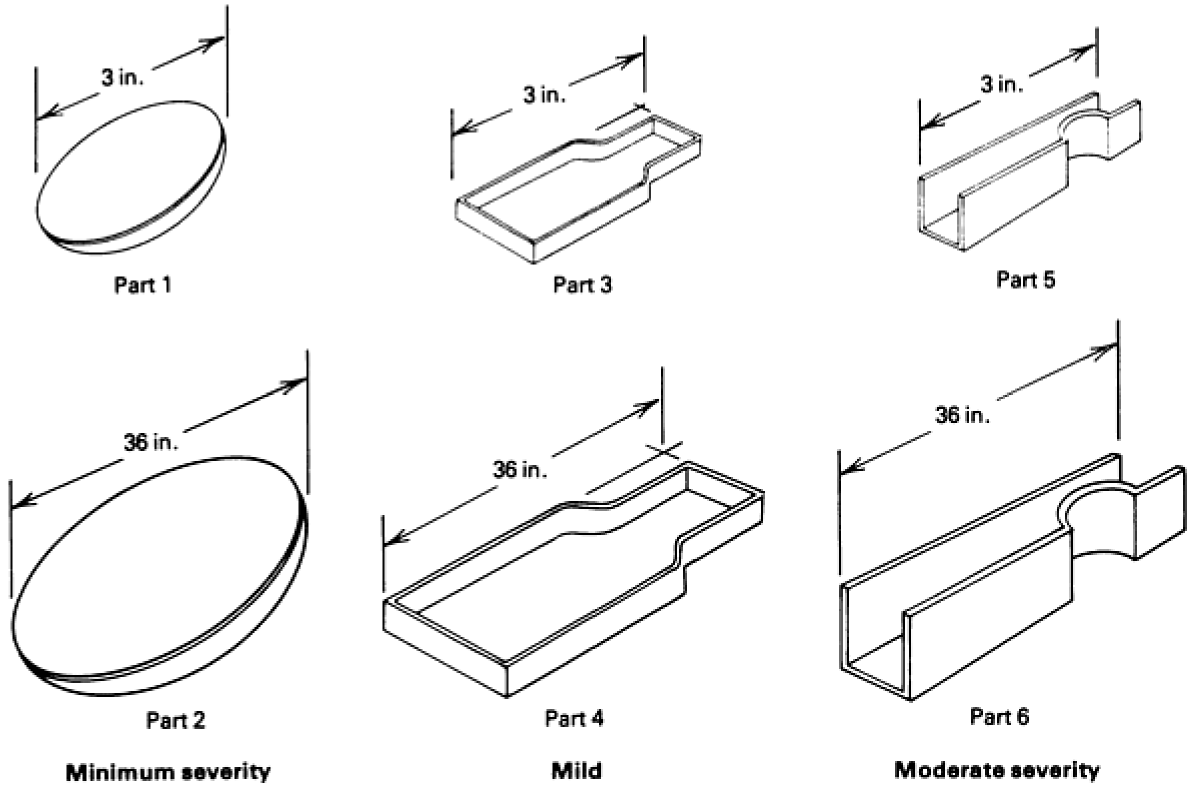


Fig. 12 Parts classified into six classes based on severity of draw. Source: Ref 36

Recommended die materials for these shapes range from plastics for low-quantity production of simple-to-moderate parts, up to the most wear-resistant (nitrided) tool steels for making severely formed parts. Parts of even greater severity, or those run in quantities larger than one million, may require dies or inserts of cemented carbide.

Effect of Part Shape. The tooling for parts 1 and 2 consists of a punch and an upper and lower die. Because there is little deformation of the sheet metal during forming of such simple parts, little or no sliding of the sheet over the lower die, and little movement over the punch, these components have very low wear rates.

Tooling for part 3 consists of a punch and lower die. In forming, the punch pushes the blank through the lower die, which results in sliding on the lower die, but there is little sliding on the punch. Therefore, the punch generally has ten times the life of the lower die made from the same material. However, wear and galling are expected in the areas of moderate shrinkage of this part, particularly when the part is formed on these single-action dies.

Tooling for part 4 consists of a punch and upper and lower die. Without the upper die, excessive wrinkling would be expected at the shrink flanges. As in part 3, a less wear-resistant material is required for the punch and upper die than for the lower die due to greater sliding occurring on the lower die.

The tooling for parts 5 and 6 consists of a punch and lower die with no upper die required because these parts are produced by stretching rather than by shrinking of the metal along the constricted region. The metal envelops the punch with minimal sliding and consequently produces about ten times more wear on the lower die than on the punch. However, the same material can be used for both the punch and lower die for part 5 because of the smaller part size and minor material cost. In the selection of die material for part 6, the critical locations are the wearing edges of the lower die. The body of the lower die could be made of cast iron with wearing edges of tool steel, and the punch could be of a material less wear resistant than tool steel, for example, low-cost alloy cast iron.

Effect of Sheet Thickness. Thick sheets of any metal will exert greater pressure on the dies than thin sheets of the same metal. Therefore, the effect of galling and wear increases as parts are made from thicker sheets, especially parts with shrink flanges. Consequently, for high-production dies working under severe wear conditions and for production of parts to close tolerances, it is often desirable to use an extremely wear-resistant material such as sintered carbide or nitrided D2 tool steel.

Effect of Tolerance Requirements. Tolerance requirements of the part affect the selection of tool material if the part is to be finished without redrawing. When the part is to be redrawn, the material used in the redrawing die is subjected to less wear than the die that performs the primary operation. A major factor in the choice between a wear-resistant material and a less costly and less wear-resistant material is the necessity for maintenance during the production run. If the production run is large and the tolerance requirements tight, costly wear-resistant materials or coated punches have to be considered.

Effect of Sheet Metal. Sheet stock of a higher hardness will usually wear dies more rapidly, but other factors, such as the presence of scale on the surface of unpickled hot-rolled steels, are sometimes of greater importance. However, scales reduce galling, which, on tool materials, may be an even more serious condition than abrasive wear and can result in frequent interruptions for reconditioning the die.

Soft brass or aluminum stock causes less wear and galling than carbon steel, whereas stainless steels and heat-resistant alloys cause more wear and galling. Possible surface treatments include chromium plating of any hardened steel, hardening of alloy cast iron, and nitriding of tool steels such as A2 and D2.

Effect of Lubrication. In low- or medium-production runs, lubricants are often used; in zinc dies, they are a necessity. However, the most effective lubricants are difficult to apply and remove. Efficient application of lubricants is particularly difficult in high-production operations where presses are being fed automatically. In such operations, die materials that are costlier but more resistant to galling (for example, aluminum bronze, nitrided D2 tool steel, and sintered carbides) should be used.

Prevention of Galling. Galling is primarily due to large stretching of sheet metal, poor tool fitting, and rough finish on the surface of tools. For short or medium runs, surface-hardened hot-rolled steel dies can be used that are less costly than tool steel dies. Tool steels are recommended for severe reductions or for forming metals that show a greater tendency to gall, such as austenitic stainless steel.

If galling is encountered, the tool fit and thickness of the stock metal should be checked first to determine whether clearances are adequate. Ironing out wrinkles causes galling. Wrinkles can be prevented by redesign of the tools.

Surface nitriding of dies made from alloy steel or alloy tool steel such as A2 or D2 minimizes galling. But in die steels containing no nitrideforming elements such as chromium or molybdenum, nitrided surfaces may spall on radii smaller than about 3 mm ($\frac{1}{8}$ in.), and plated surfaces may spall on radii less than 6 mm ($\frac{1}{4}$ in.). Hard chromium plating will usually eliminate galling of mild steel, alloy steel, and tool steel dies. But to avoid cracking of the hard coating in operations involving high local pressures, hardened steels such as alloy or tool steels should be used as the substrate. Dirt, grit, and shot fragments on the sheet cause greater damage to nitrided and chromium-plated tools than to hardened tool steels; hard particles may cause minute spalling and small pits, especially at radii in areas of high forming pressure.

Galling is less likely if the die and stock materials are dissimilar in hardness, chemical composition, and surface characteristics (for example, aluminum bronze tools for forming carbon steel and stainless tools for aluminum parts).

Galling is also related to the thermal softening resistance of the die material. Galling and pickup are caused by the extremely high frictional heat generated between the die material and the steel being formed, which results in the softening of the die material and increased pickup.

On the basis of tests conducted by an automobile manufacturer, it was concluded that (Ref 36):

- Metal pickup on the faces of flanging dies is due to the frictional heat generated by two metals rubbing together under high pressure
- Excessive heat is generated if the length of contact is longer than needed to form the flange
- Correct die lubricants should be chosen. For example, when heavy gages are being formed, an extreme-pressure (EP) die lubricant should be used

Wear and galling are less severe with soft stock material, such as aluminum and copper alloys, than with low-carbon steel; they are more severe with high-strength metals such as stainless steels and heat-resistant alloys.

Die Wear and Die Life Studies. The important factors influencing the wear and life of sheet metal forming dies are the composition and hardness of the die material and the part, and the thickness, shape, and quantity of the parts to be formed. Effects of these factors are shown graphically in Fig. 13. The data for Type 4140 steel show the effect of die hardness: its die life increases about 800% when nitrided (Fig. 13a). Effects of the tool material composition and hardness are seen in the forming of the label holder (Fig. 13b), and in the forming of the bushing (Fig. 13c). M4 and D7 at 65 HRC gave more than twice the die life of D2 (61 HRC) in flanging the shredder ring of 17-7 PH stainless steel shown in Fig. 13(d).

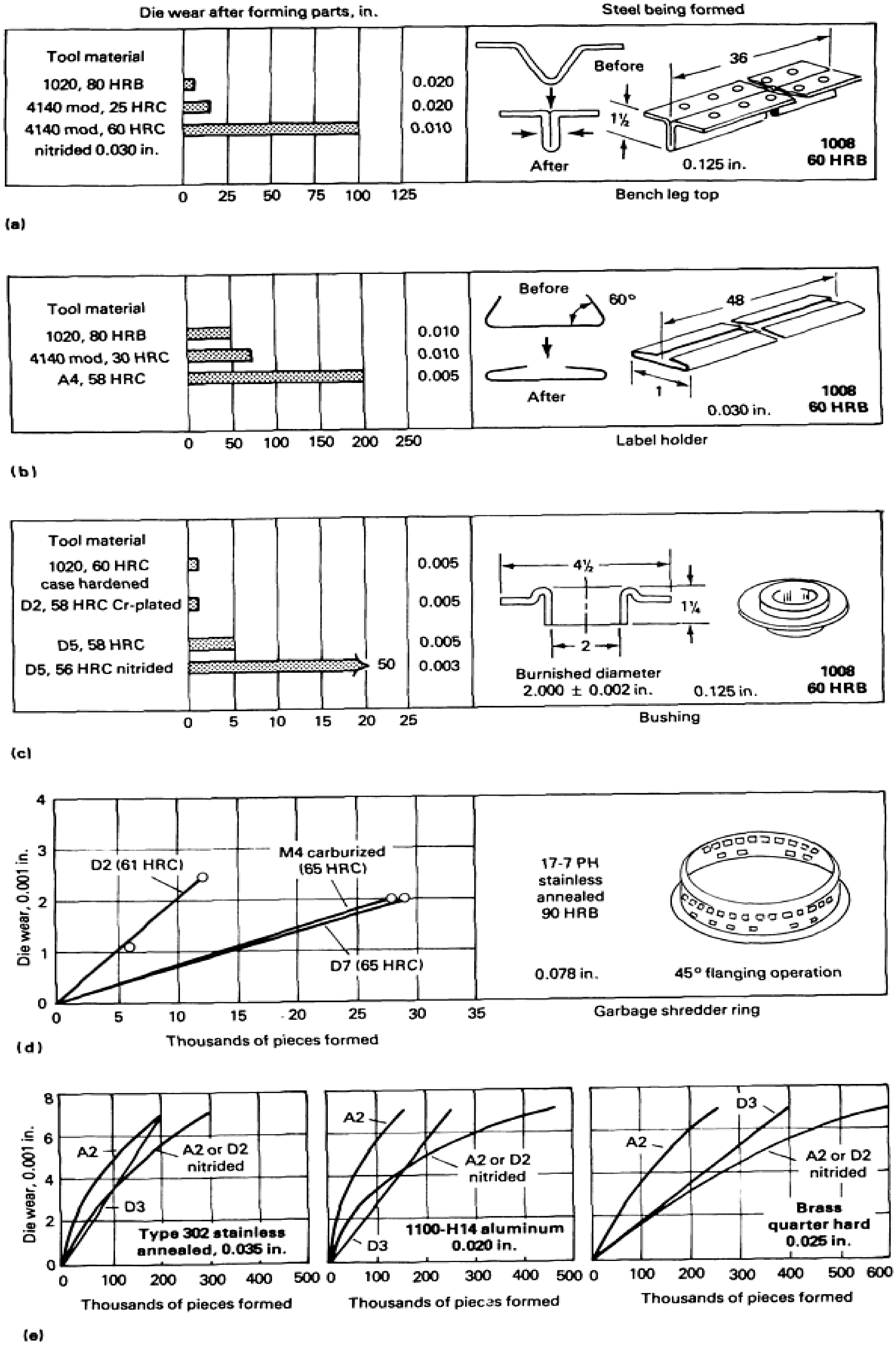


Fig. 13 Data on die wear and die life. The three plots in the bottom row relate to small instrument parts having a maximum area of 19 cm^2 (3 in.^{2}). Source: Ref 36

The plots in Fig. 13(e) show the influence of die and stock material on wear of dies for making parts of similar shape. The wear rate gradually decreases for nitrided steels due to the wearing away of the outer layer (white layer).

Wear in Deep-Drawing Dies

This section deals with the wear of material for dies to draw round and square cup-shaped metal parts in a press, primarily by conventional drawing in which each reduction is made in the same direction. Deep drawing is associated with extension and bending of the sheet over the punch nose, together with the gathering and compression of the sheet metal as it is drawn through the die. For economy of manufacture, the drawn part is always produced in the least number of steps using the least number of dies possible. Ironing is used almost universally in multi-operation drawing and to increase uniformity of wall thickness. Each operation is designed for maximum practical reduction of the metal being drawn; the examples given in this section involve maximum reductions of about 35%.

The performance of a deep-drawing die is determined by the amount of wear or galling during a production run. Wear of a given die is determined largely by the material and thickness of the sheet steel, sharpness of die radii, lubrication, die design, and finish. The amount of wear on die radii can vary by as much as a factor of 20 between the sharpest and most liberal radii.

In drawing square cups, the formation of wrinkles at the corners accompanied by high localized pressures may produce prohibitively high rates of wear. Small corner radii will result in greater wear and shorter life of the dies.

High localized pressures caused by inaccurate fitting of dies during tryout, rough surface on the drawn sheet or on the working surfaces of the die, inadequate lubrication, and poor maintenance (stoning) of dies are typical of uneconomical practice.

Effect of the Thickness of Sheet Metal. In drawing thick sheets of a given metal, the pressure on the dies increases in proportion to the square of the sheet thickness. The die pressure is higher on the draw radius, and increasing sheet thickness will localize the wear in this high-pressure area without a similar effect on other surfaces of the die.

Thick stock will wrinkle less than thin stock; therefore, the pressures required to prevent wrinkling are less than those required for thin stock. However, heavy sheets are often drawn without a pressure pad. Therefore, the wear-resistance requirements for pressure pads used on heavy stock are no greater than those for thin material.

Reference 36 gives recommendations for tool materials for drawing and ironing sheet steel and plate ranging in thickness from 0.38 to 13 mm (0.015 to 0.500 in.). Alloy cast iron is often flame hardened for long runs and thick sheets; for small quantities and thin sheets, it may be used as-cast. For the mildest applications, unalloyed gray iron may be used.

Prevention of Galling. Common causes of galling are stretching sheet metal beyond practical limits, tool fitting with poor alignment or insufficient die clearance for the given sheet thickness, wrinkles, the use of galling-susceptible tool steel, and rough finish on the surface of tools.

Die materials to resist galling can be selected on the basis of their resistance to softening under heat. For example, T15 steel is the die material that has highest resistance to softening and thus the best galling resistance; W1 steel has the poorest resistance to softening and galling.

For parts drawn of carbon steel or nonferrous alloy sheet, the die material can be selected without regard to galling; as a finishing operation, the punch and die should be either nitrided or chromium plated. If chromium-and molybdenum-containing tool steels such as A2, D2, D3, or D4 have been selected, the smoothly ground tools should be nitrided and polished or buffed after nitriding. Otherwise, the tools should be hardened to at least 60 HRC, smoothly finished on the wearing surfaces, hard chromium plated, and the plating polished or buffed. Punches should be plated to a thickness of 0.05 to 0.01 mm (0.0002 to 0.0004 in.). To prevent spalling or flaking of the plate, dies should not be plated more than 0.005 mm (0.0002 in.) thick.

When the parts are drawn from stainless steel or from high-nickel alloy steel, the draw ring material with best resistance to galling is aluminum bronze. The second choice is either D2, D3, or D4 (smoothly ground, nitrided, and polished). The third choice is alloy cast iron (quenched and tempered to 400 to 420 HB).

Punch material is selected without regard to galling and is then chromium plated (as with die materials), unless cast iron is chosen.

Sintered carbide, which is only as efficient as the lubricants used, has proved economical for nonferrous alloys, carbon steel, and stainless steel in many long, continuous runs; however, galling will occur as soon as lubrication becomes faulty.

Effect of Lubrication. Correct lubrication of the parts being drawn is essential to reduce friction, wear, and galling. In fact, deep drawing is impossible without lubrication. In actual practice, die materials are selected after trials employing the production lubricants. If excessive wear or galling occurs, a better lubricant is usually applied. For extremely difficult draws, the best lubricants are usually applied at the outset.

Chromium Plating to Reduce Wear. Chromium plating is used on tool steel draw rings to improve life. On punches, its primary function is to reduce the frictional forces and facilitate removal of the parts from the punch after the sidewalls have been ironed tight to the punch. Usually chromium plating improves punch life somewhat less than it would be improved by the use of the next best tool steel.

For successful performance of tools, chromium plating must always be deposited on a surface harder than 50 HRC, and the thickness of the coating should preferably be 0.05 to 0.01 mm (0.0002 to 0.0004 in.) and never less than 0.0025 mm (0.0001 in.). This gives the required hardness and reduction of friction without excessive spalling or chipping at corners. Chromium-plated dies should be heated to 150 to 205 °C (300 to 400 °F) for a minimum of three hours immediately after plating to minimize the possibility of hydrogen embrittlement.

Die Wear in Hot Forging Dies

In forging steels, die life is often controlled by abrasive wear. Thus, die wear and die life are often thought to be synonymous. However, wear is but one of the several mechanisms by which dies are rendered unusable (Ref 40, 41). Another common mechanism in hot forging is thermal fatigue, or thermal cycling, which gives rise to superficial cracks often known as heat checks. Analogous with thermally induced cracking is mechanical fatigue, or cracking that results from the cyclic application of the forging loads. If the loads are very high or the dies relatively soft, plastic deformation of the dies may occur, making it impossible to impart the desired shape to the workpiece. Although it is not unusual for several of these mechanisms to contribute to die failure, abrasive wear will be emphasized in this section.

Methods of Characterizing Abrasive Wear

The amount of die material removed because of abrasive wear is directly proportional to the interface pressure and the amount of relative sliding, and inversely proportional to the hardness of the metal surface. Most forging dies are typically of rather complex geometry. Therefore, the interface pressure and amount of relative sliding can vary from one area to another. Hence, characterization of abrasive wear in a systematic manner typically makes use of simulative tests of simple geometry.

Most simple die wear studies have made use of upset compression of cylindrical billets (Ref 42, 43, 44, 45, 46). In these tests, measurements of die wear were obtained directly by measuring the surface roughness before and after the tests. Upsetting of cylinders on flat dies may be a convenient way for forging a large number of specimens for the purpose of testing a die steel. While this test may accurately reflect die wear characteristics in some cases, it differs from closed-die forging in some important respects:

- Because the flat-bottomed specimen is placed on a flat die prior to forging, a substantial amount of heat can be conducted from the specimen to the die. Thus, at that interface, the specimen is colder and the die is hotter than would be the case in an actual forging die where the irregularly shaped die makes line and point contact with the billet
- In upsetting a cylinder, the metal flow is all lateral, or in a direction perpendicular to the ram motion. In an actual forging, metal flow will occur in both lateral and longitudinal directions; in some cases, longitudinal flow can be very rapid, exceeding by far the velocity of the press ram. In these cases, the sliding velocity at the interface can be an important factor contributing to die wear

Thus, it must be concluded that only die wear studies conducted with actual forging dies can give reliable results.

Most evaluations of die wear in impression die have made use of dies of simple geometry. These include the investigations of Silva and Dean (Ref 47) and those of Doege, Melching, and Kowallick (Ref 48). The latter workers used an automated heading machine to forge 1045 steel disks into cups via an extrusion-type process. Die heaters were employed to keep the temperature constant. After heading 300 pieces, a surface analyzer was used to measure the wear of the dies, and the average distance between the worn and unworn profile was taken as the measure of wear.

A very extensive series of die wear studies was conducted by Netthöfel (Ref 49), who used the axisymmetric die shown in Fig. 14. Originally, Netthöfel designed and tested an axisymmetric die made of a number of pie-shaped segments. Each segment could be made of a different material and, thus, it would be possible to investigate the wear of several die materials with a single series of forging experiments. However, this approach was not successful. During forging, the segments separated and the forged materials penetrated into the spaces, or "cracks," between the adjacent segments. Consequently, it was decided to use pin-shaped inserts made from the die materials to be investigated. Each pin had a 16 mm (0.63 in.) diameter and was 0.04 mm (0.0016 in.) larger than the hole into which it was placed. The pins were shrink-fitted into the various holes located in the axisymmetric die, as shown in Fig. 14.

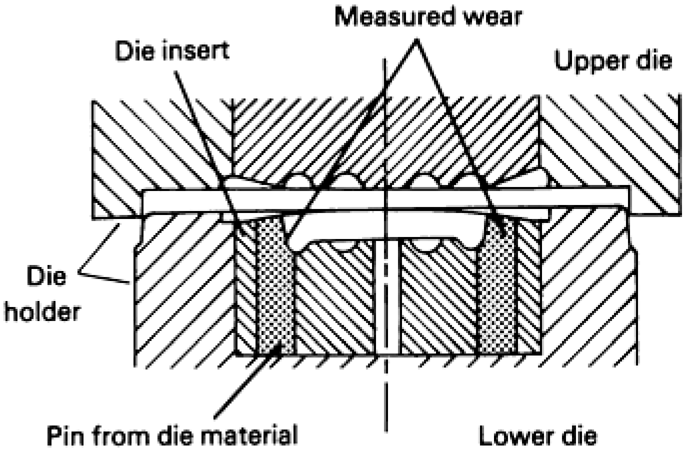


Fig. 14 Cross section of a closed-die forging test setup. Source: Ref 49

Netthöfel used a friction screw press in his studies, and operated it at an initial forging speed of 0.5 m/s (20 in./s). In conducting most of the forging trials, a maximum number of 6000 parts was forged from 0.53% carbon steel (0.54% C, 0.26% Si, 0.7% Mn, 0.02% P, 0.02% S). In order to maintain all of the conditions at approximately constant levels, the die temperatures were controlled by heating the dies with a gas flame and the part was ejected from the lower die using an ejector. The forging cycle was approximately 13 s. The round-cornered square billets were heated inductively to 1200 °C (2190 °F), and flattened in a mechanical press, prior to forging in the screw press with the dies seen in Fig. 14. As a lubricant, Netthöfel used sawdust, which was sprayed over the lower dies and the workpiece material before each blow. Die wear was measured (at the locations indicated in Fig. 14) by taking a cast of the worn die after a number of forgings. In preliminary experiments, not only dies with pin inserts, but also those from full materials were investigated. The results indicated that die wear, as measured in using pin inserts, was identical to that obtained in full material dies under identical forging conditions.

Factors Affecting Abrasive Wear

Die Material. Of all the factors that influence abrasive wear, the one easiest to understand and quantify is that of die material and hardness. In general, increasing alloying content and die hardness both tend to increase the resistance of forging die steels to abrasive wear. As was discussed earlier in this article, low-alloy die steels such as 6F2, 6G, and 6H1 generally have poor resistance to wear as compared to hot-work die steels such as H13 and H26 (Table 1). This is because the microstructures of the latter steels are not only inherently more resistant to wear, but they also tend to be more stable at higher temperatures.

The effect of various alloying elements on wear has been discussed by various authors. Kannappan (Ref 40) summarized results on the wear resistance of several alloy steels and concluded that it increases with increasing contents of carbon and carbide-forming elements. Further, he surmised that the presence of noncarbide-forming elements in martensitic die steels

may even be detrimental. Of the carbide-forming elements, greater wear resistance is developed in the order of chromium, tungsten, molybdenum, and vanadium, with the effectiveness in reducing wear being in the ratio of 2:5:10:40, respectively, when the die temperature is between 250 and 550 °C (480 and 1020 °F). Thus, vanadium and its associated carbides are eight times as effective as tungsten and its carbides in reducing wear. Thomas (Ref 43) made a similar ranking of tungsten, molybdenum, and vanadium, with relative effectiveness in reducing wear being in the ratio of 10:20:40.

The work of Aston, Hopkins, and Kirkham (Ref 46) provides a more detailed insight into the effect of alloying on wear of forging die steels. Their results on eleven die steels tempered to nearly the same hardness (Fig. 15 and Tables 8(a) and 8(b)) were summarized as follows:

- The steel with the lowest alloy content (number 1) has the least wear resistance. It is useful primarily in hammer applications because of its low price, good toughness, and ease of machining
- Steels of moderate wear resistance (numbers 2, 3, 4, 5, and 9) have large amounts of chromium (1.25 to 12%) and molybdenum (0.5 to 1.5%) and some vanadium (0.07 to 1.5%)
- Steels of highest wear resistance (numbers 6, 7, 8, 10, and 5) include those with the highest alloy content. However, several of these die steels (notably, numbers 7 and 10) have only moderate alloying

Table 8(a) Nominal composition of steels tested by Aston *et al.* (Ref 46)

See also Fig. 15.

Steel No.	Composition, %								
	C	**Cr**	**Mo**	**W**	**V**	**Ni**	**Mn**	**Nb**	**Si**
1	0.55	0.75	0.25	. . .	. . .	1.5	. . .	. . .	. . .
2	0.34	1.4	0.6	3	0.6	5.6	. . .	. . .	. . .
3	0.45	1.25	0.5	. . .	0.07	1	1	. . .	. . .
4	0.2	12	0.9	. . .	0.3	. . .	. . .	. . .	. . .
5	0.35	5	1.5	1.5	1.5	. . .	. . .	. . .	. . .
6	0.2	12	1.5	. . .	0.3	2.5	. . .	0.7	. . .
7	0.4	3.25	1	. . .	0.25	0.3	. . .	. . .	. . .
8	0.34	13	. . .	3	0.6	. . .	. . .	. . .	2
9	0.35	5	1.5	. . .	1	. . .	. . .	. . .	. . .
10	. . .	5	2	. . .	0.4	. . .	. . .	. . .	. . .

Table 8(b) Heat treatments of die steels tested by Aston *et al.* (Ref 46)

See also Fig. 15.

Steel No.	Hardening temperatures		Quench medium	Temperature		Time, min	Hardness, HV
	°C	**°F**		**°C**	**°F**		
1	830	1525	Oil	500	930	45	399
2	1000	1830	Oil	600	1110	45	432
3	860	1580	Oil	660	1220	40	409
4	1050	1920	Air	590	1095	60	394
5	1040	1905	Air	660	1220	30	401
6	1050	1920	Air	590	1095	100	408
7	910	1670	Oil	650	1200	60	398
8	1120	2050	Oil	880	1615	45	427
9	1040	1905	Air	670	1240	45	408
10	1040	1905	Air	670	1240	45	406
11	1020	1870	Furnace	550	1020	30	405

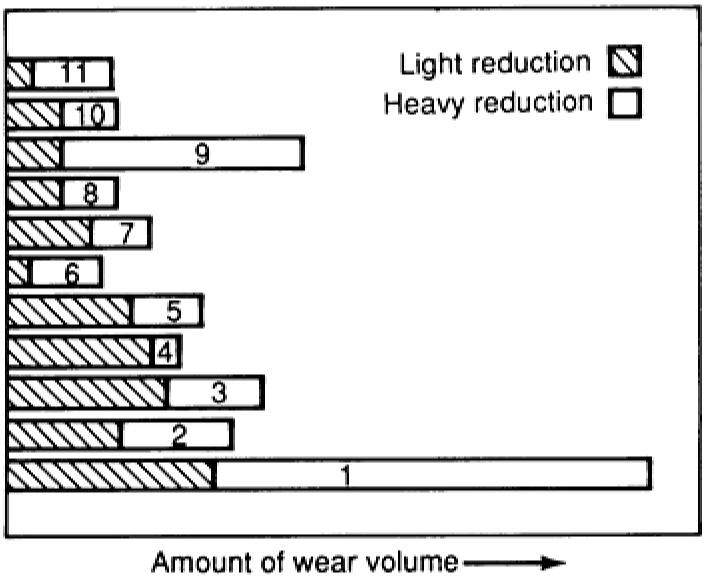


Fig. 15 Results of tests on the effects of alloying on wear of forging die steels. See Tables 8(a) and 8(b) for compositions and heat treatments of steels. Source: Ref 46

From these results, they concluded that good wear resistance is obtained when the total alloy content is in excess of 3%. They remarked on the particularly strong effect of molybdenum on reducing wear (Fig. 16) but noted that quantities in excess of 2% were not needed.

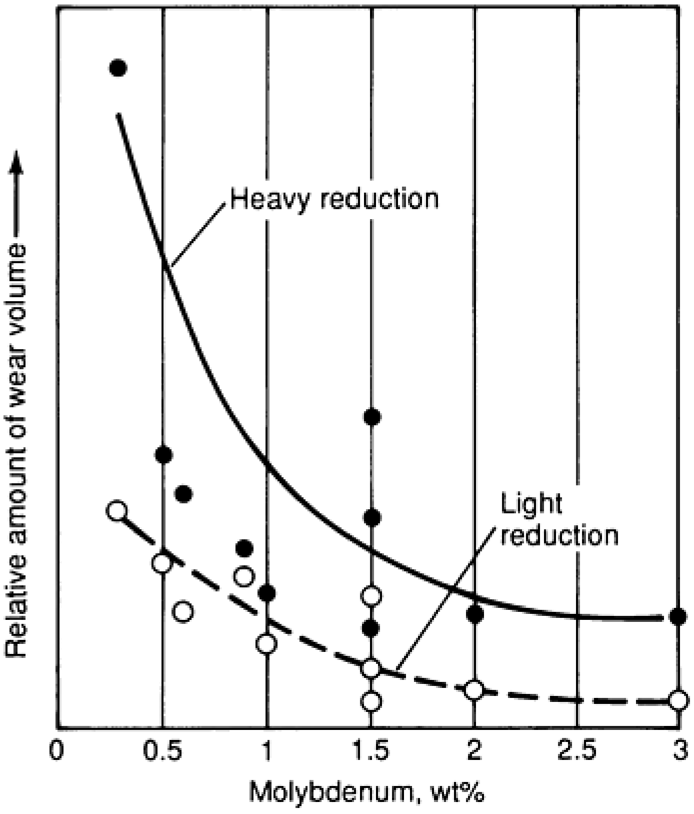


Fig. 16 Effect of molybdenum content on wear resistance of die steels. Data are a cross-plot of results shown in Fig. 15. Source: Ref 46

The results by Thomas (Ref 43) also showed a strong correlation between wear resistance and alloy content (Fig. 17). The low-alloy steel (No. 5 die steel, which is equivalent to steel number 1 in Fig. 15, and 8(a) and Table 8(b)) had relatively poor wear resistance when compared to the 5% Cr hot-work die steel (H12) and the 12% Cr steel. This was true irrespective of the hardness level to which the steels were tempered and even, to a certain degree, of the type of workpiece material.

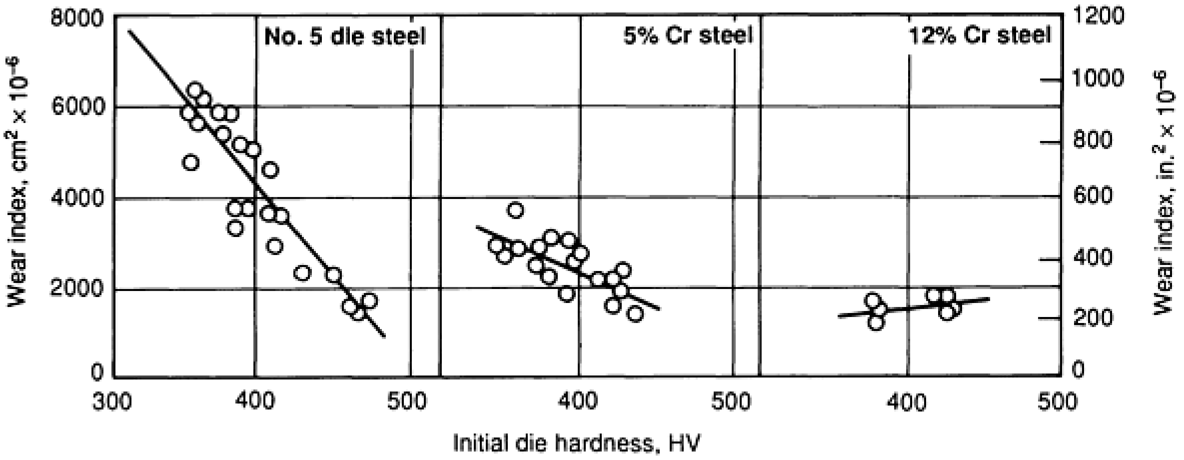


Fig. 17 Influence of initial die hardness on wear of die steels. The wear index is defined as average cross-sectional area of wear depressions in dies. No. 5 die steel: 0.6 C, 0.3 Si, 0.6 Mn, 1.5 Ni, 0.6 Cr, 0.25 Mo; 5% Cr steel: 0.33 C, 0.3 Si, 1.0 Mn, 5.0 Cr, 1.5 Mo, 1.5 W, 0.5 V; 12% Cr steel: 0.1 C, 0.25 Si, 0.7 Mn, 2.4 Ni, 12.0 Cr, 1.8 Mo, 0.35 V. Source: Ref 43

These findings have been verified and further expanded by other researchers. For example, in their carefully controlled forging experiments, Doege, Melching, and Kowallick (Ref 48) and Hecht and Hiller (Ref 50) found low-alloy steels to have far inferior wear resistance as compared to hot-work die steels (Fig. 18) because alloying led to higher hardness and the ability to retain strength at high die temperatures. The work of Netthöfel (Ref 49) is also in agreement with these observations.

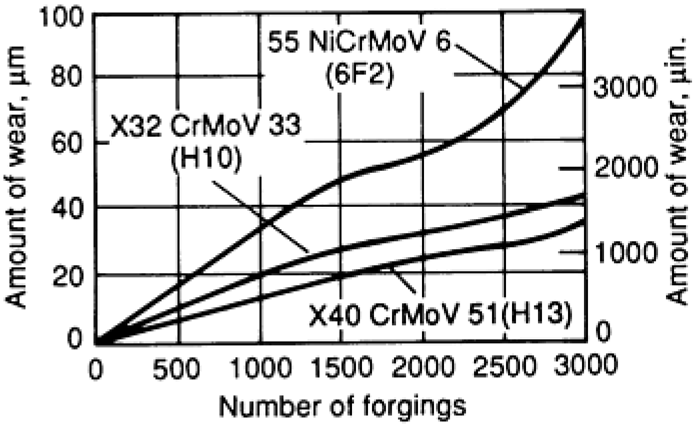


Fig. 18 Amount of wear of hot-work tool steels as a function of the number of forgings. Equivalent steels are in parentheses. Source: Ref 48

Up to now, most of the discussion of alloying has centered on die steels. Several workers have also investigated the wear characteristics of nonferrous die materials. In Netthöfel's (Ref 49) experiments on forging die wear, it was found that the nickel-base alloy Nimonic 90 had a wear resistance between that of an H12 and H19 steel at a die temperature of 255 °C (490 °F). This is an important finding in view of the fact that the nickel-based alloys are generally many times the cost of

the die steel alloys and are also harder to machine. Notthöfel's finding was verified to a certain extent by Ali, Rooks, and Tobias (Ref 42) in their die wear studies in a high-energy-rate forming (HERF) machine (Fig. 19). Although the die were depended on whether the top die or bottom die was examined, it was found that Nimonic 90 was only slightly better than a steel similar to H19 (WEX). Thus, the results reported in Ref 42 and 49 point out the fact that the nickel-base die materials should be reserved for hot-die and isothermal forging applications for which die steels are inappropriate.

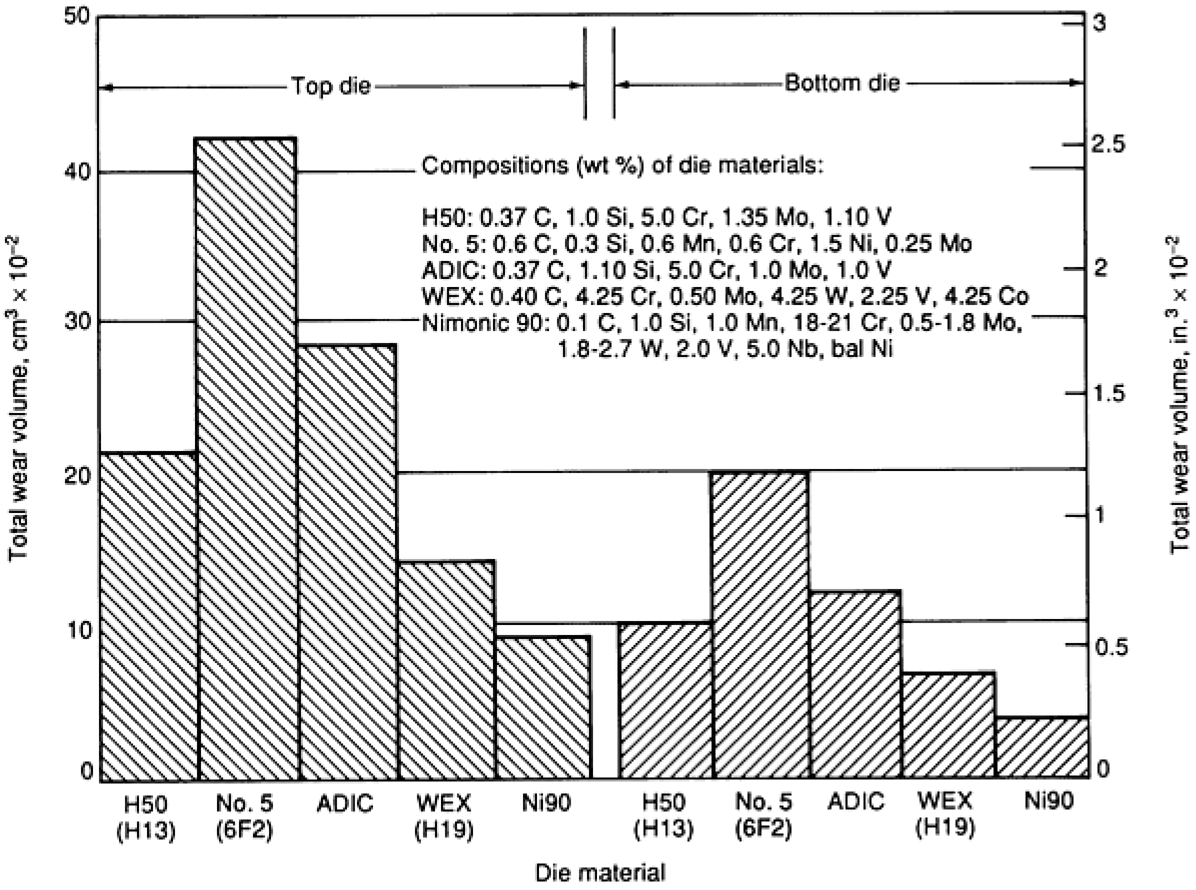


Fig. 19 Total wear volumes for die materials at a mean hardness of 44 HRC. Source: Ref 42

Die Hardness. Die hardness is another factor whose influence on abrasive wear is easy to quantify. The effect of die hardness is best realized through an understanding of the die wear process itself. Misra and Finnie (Ref 51) have summarized a large amount of work on abrasive wear and concluded that two basic processes are involved. The first is the formation of plastically deformed grooves that do not involve metal removal, and the second consists of removal of metal in the form of microscopic chips. Because chip formation, as in metal cutting, takes place through a shear process, increased metal hardness could be expected to diminish the amount of metal removal via abrasive wear. This trend is exactly what has been observed.

The effect of hardness on wear is seen in data on a variety of steels quoted by Kannappan (Ref 40) and by Thomas (Ref 43), which have been discussed previously. From examination of Fig. 17, it is apparent that the dependence of wear rate on hardness is greatest for low-alloy die steels such as 6F2 (No. 5 die steel in Fig. 17). Such a trend has also been reported by Kannappan (Ref 40) in data on several low-alloy and hot-work die steels.

Kannappan (Ref 40) has also discussed the correlation between hardness and wear of die steels with microstructures different from the typical die steel structure of tempered martensite. It has been found that the isothermal heat treatment of steels to produce lower bainite results in better wear resistance (Fig. 20). Supposedly, this effect is a result of the fact that isothermal transformation/hardening causes fewer stresses and microscopic cracks (which promote abrasive failure) than does a thermal martensitic transformation.

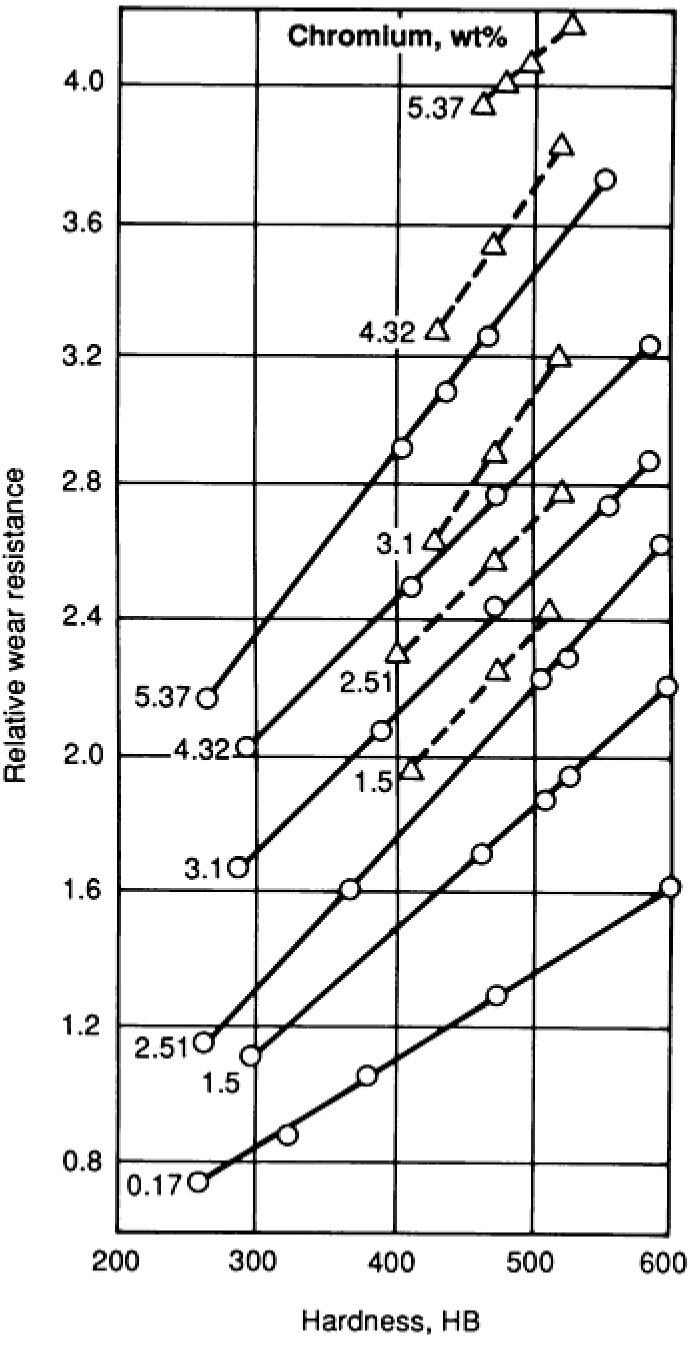


Fig. 20 Relative wear resistance with respect to hardness of selected chromium steels with 0.55% C. Note the difference between the effect of quenching followed by tempering (solid lines) and the effect of isothermal treatment/quenching to a lower bainitic region (dashed lines). Relative wear resistance is defined as a number directly proportional to the applied interface pressure and the amount of relative sliding and inversely proportional to the total wear volume. Source: Ref 40

Workpiece Temperature. Several researchers have commented on the effect of workpiece temperature on die wear. In his investigation of wear of hammer dies, Thomas (Ref 52) found that in forging of steels, wear increased at first with billet temperature up to 1100 °C (2010 °F) and then decreased with increasing temperature (Fig. 21). The initial increase can probably be attributed to the increase in the amount of scale on the billets, which acts as an abrasive during the die wear process. However, above 1100 °C (2010 °F), the flow stress drops off rapidly enough to minimize the interface pressure during forging and therefore decrease the effect of scale. A similar finding was made by Doege, Melching, and Kowallick (Ref 48), who attributed an increase in die wear as the billet temperature was raised from 800 °C (1470 °F) to 1100 °C (2010 °F) to an increase in the die surface temperature and a simultaneous decrease in wear resistance.

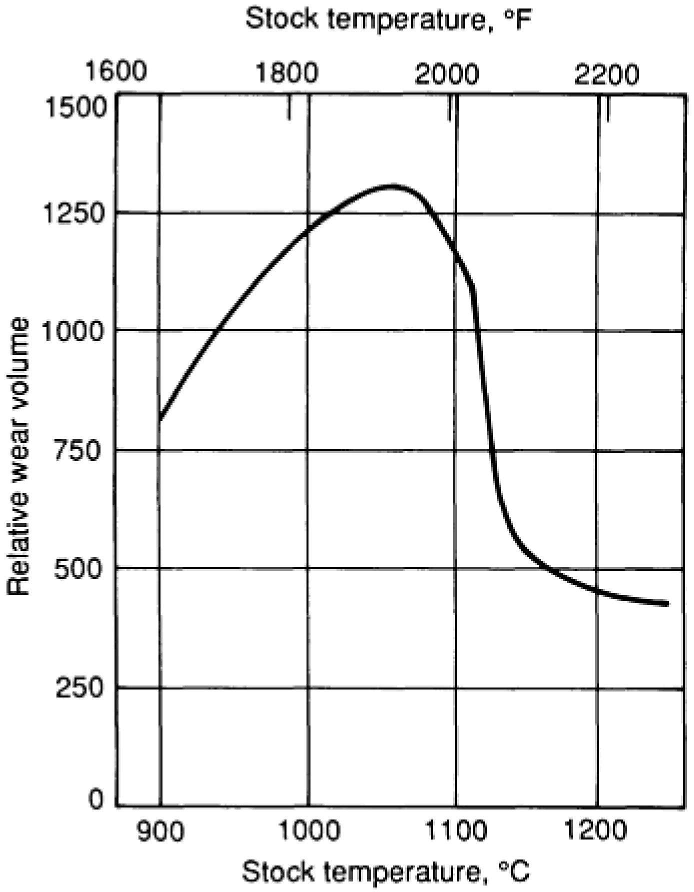


Fig. 21 Effect of workpiece temperature on wear. Source: Ref 52

Lubrication/Die Temperatures. The effects of lubrication and die temperature on die wear have been interpreted in a variety of often-conflicting ways in the literature. This is because lubricants and die temperature influence: lubricity, and hence the amount of metal sliding during forging; the interface pressure during deformation; and the heat transfer characteristics between the dies and workpiece during conventional hot forging. The last item is important not only through its influence on heat absorption into the dies, and thus thermal softening and decreased wear resistance of the dies, but also through its effect on the performance of the die and billet lubricants themselves.

Investigations into the effect of lubrication on die wear in simple upsetting have shown that wear is greatly increased when the dies are lubricated versus when they are not. This effect is shown in the results of Singh, Rooks, and Tobias (Ref 44) from upset tests in a HERF machine (Fig. 22). The same phenomenon has been demonstrated by Thomas (Ref 43), who upset successive lots of 1000 samples each on a flat die in a mechanical press. In these tests, the amount of wear was greater for the lot involving lubricated compression tests (Fig. 23). From these findings, one might conclude that wear increases with lubrication because of increased sliding and that lubrication is detrimental in forging. Thomas clarified this point, however, by calculating the amount of wear for equivalent amounts of metal flow past a given point; he found that lubrication reduces wear by a factor of 3 when compared to forging without lubrication. Moreover, he emphasized that in closed-die forging, the amount of metal sliding is fixed by die and preform design and not lubrication. Thus, the amount of sliding over the flash land, where wear is usually greatest, depends on the amount of flash that must be thrown and not on the efficiency of the lubricant employed. Because the amount of flash will be roughly the same with or without lubrication, employing lubricants in closed-die forging should reduce abrasive wear of the flash land and other parts of the die cavity.

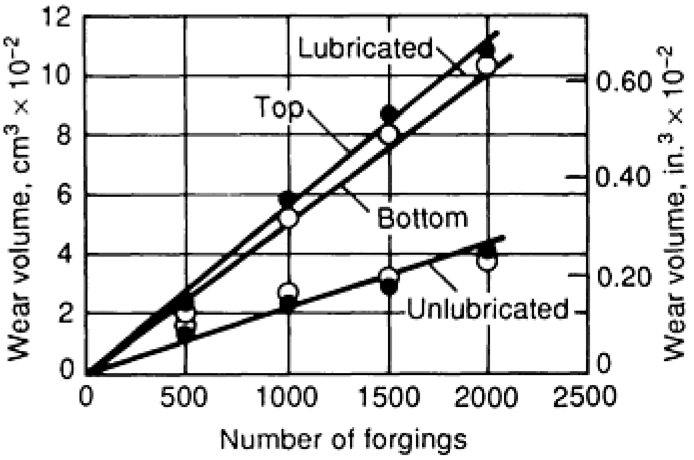


Fig. 22 Effect of lubrication on forging die wear. Source: Ref 44

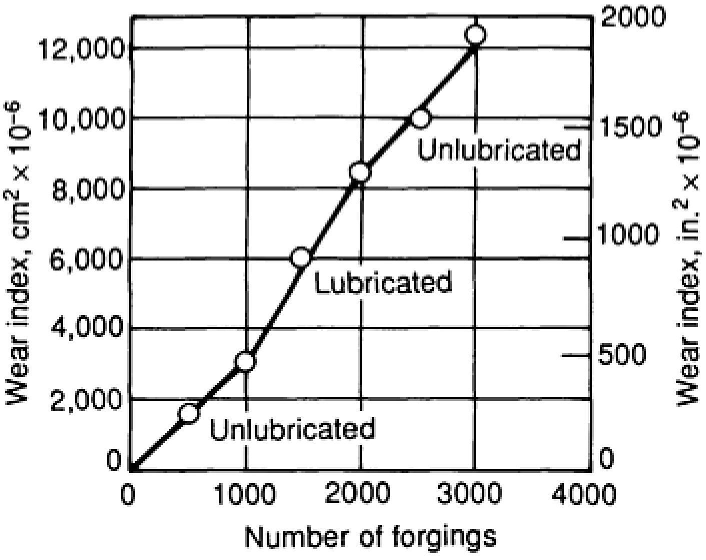


Fig. 23 Effect of lubrication on forging die wear. Wear index is defined as the average cross-sectional area of wear depressions in the die. Source: Ref 43

The interaction of lubrication and die temperature effects was demonstrated by Rooks (Ref 45) in upset tests on a HERF machine. These tests were run with various bulk die temperatures, dwell times, and cycle times. Dwell time in the HERF operation includes a short forging phase, a somewhat longer "bouncing" phase, and an extended "after-forging" phase during which the dies and billet are in contact under low pressure. Results established that die wear after upsetting of 1000 billets decreased with increasing die temperature. This was correlated with decreased amounts of sliding at higher die temperatures due to an increase in the coefficient of friction.

The effects of dwell time and cycle time on die wear were also examined by Rooks (Ref 45). Increasing dwell time increases die chilling. As a result, metal flow is hindered and die wear is reduced. Increased cycle time (time between forgings) tends to have the reverse effect of increasing dwell time (that is, it increases die wear because of lower coefficients of friction and more sliding). However, these effects have been found to be very slight in upset tests, conducted in a HERF machine (Ref 45).

A striking die wear feature that Rooks (Ref 45) and Ali, Rooks, and Tobias (Ref 43) noted concerns the generally higher wear experienced by the top die versus the lower die, which is most noticeable in their lubricated upset tests (Fig. 19). This can be attributed to greater chilling on the bottom die because the hot workpiece was placed on it prior to forging.

This could, therefore, have been expected to lead to greater friction, less sliding, and thus less abrasive wear than the top die experienced.

From a practical standpoint, increased production rate in a forge shop may be expected to lead to lower die life. This is almost certainly a result of increased die temperature. In forging under production conditions, the die surface temperature observed between two consecutive forging blows seems to remain unchanged throughout a production run (Ref 49). During the actual forging operation, the die surface temperatures increase and reach a maximum peak value and decrease again when the dies are separated and the forging is removed. In case the forging "sticks" in one of the dies, the peak surface temperature of that die may increase further and contribute to die wear. Therefore, in conducting die wear studies, it is suggested that an ejector be used to remove the part after forging, so that die temperatures do not increase because a forging sticks in the die. In forging of steel at 1200 °C (2190 °F) with dies at about 250 °C (480 °F), surface temperatures will reach approximately 750 °C (1380 °F) if perfect and ideal contact occurs between the forging and the die. In reality, however, due to scale and oxidation at the die/material interface, the peak surface temperatures during forging reach 500 to 600 °C (930 to 1110 °F) in mechanical presses and 650 to 700 °C (1200 to 1290 °F) in hammers. As an example, die temperatures obtained by Vigor and Hornaday (Ref 53), in forging steel in a mechanical press are given in Fig. 24. It can be seen that the temperature gradient is very large at the vicinity of the die/material interface.

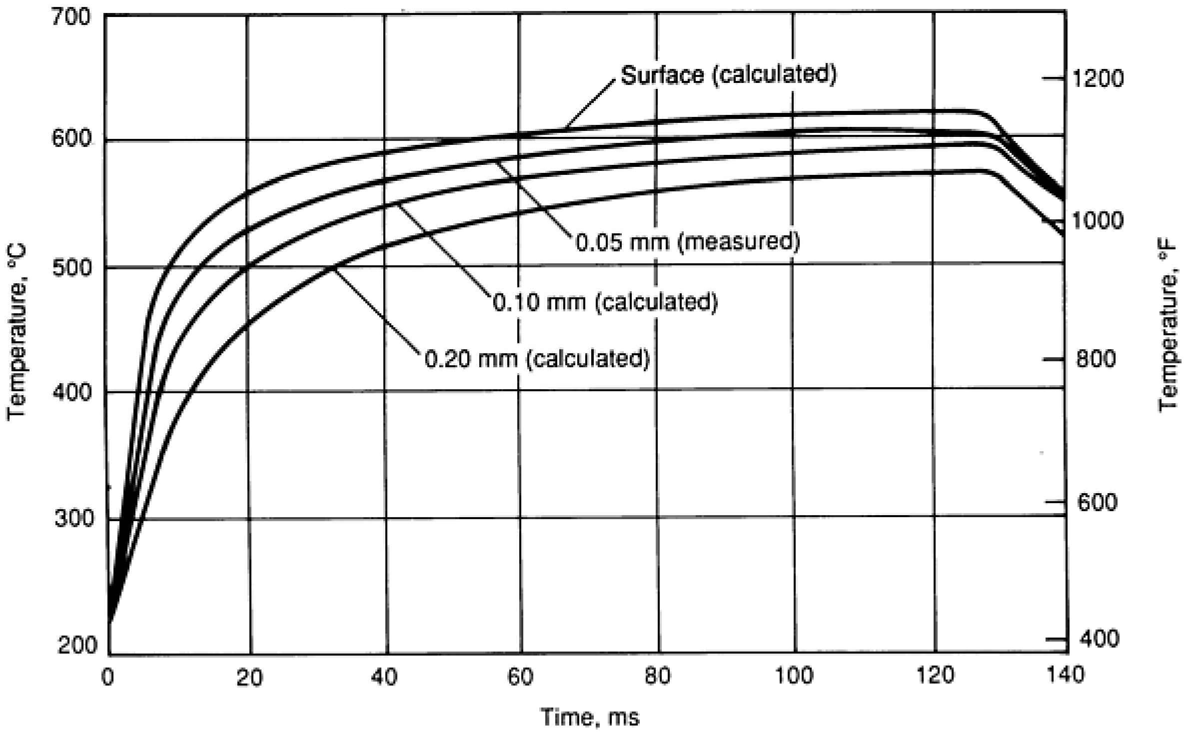


Fig. 24 Temperatures at the surface and at various depths in forging dies obtained during forging 1040 steel without lubricant. Source: Ref 53

The effects of sliding on die wear are also qualitatively well known in forging practice. These effects are taken into account in designing preforms to ensure that more "squeezing" and less lateral flow and sliding action take place during finish forging.

Methods of Improving Resistance to Abrasive Wear

From the discussion of the factors that influence abrasive wear, one can deduce methods to improve die performance controlled by this failure mechanism. Perhaps the most direct method is to employ a die steel that is more resistant to wear, that is, one that is harder and that retains its hardness at high die temperatures (Ref 54). This could mean changing from a low-alloy die steel to a chromium hot-work die steel. The decision to make such a change should be based on the suitability of the new die steel itself in the forging operation and the trade-off between expected increases in die life and increases in material (and machining) costs.

Coating, hardfacing, and surface treatment of forging dies often can be employed to improve wear resistance as well. Information regarding specific coating and hardfacing alloys (and the methods of their application) and surface treatments such as nitriding and boriding is contained in the following section of this article and will not be reviewed here. However, there are numerous instances of such methods increasing die life. These include the use of chromium and cobalt-base coatings (Ref 48, 55), weld deposits of higher-alloy steels onto low-alloy steels (Ref 56), weld deposits of nickel and cobalt hardfacing alloys on die steels (Ref 57, 58, 59), ceramic coatings (Ref 60, 61), and surface nitriding (Ref 54, 62, 63).

Another means of reducing wear in the forging of steel involves reducing the scale on heated billets; scale acts as an abrasive during the sliding that occurs between the dies and workpiece. Thomas (Ref 52) estimates that poor control of scale can reduce die life as much as 200%. Methods of reducing scale are relatively obvious and include the following:

- Using a reducing, or inert, furnace atmosphere
- Using a billet coating to prevent oxidation
- Minimizing time at temperature in the furnace or using induction heating

One final means of decreasing the problem of wear is through improved redesign of the blocker shape. This is an important consideration because wear is strongly dependent on the amount of sliding that occurs on a die surface. Thus, it is possible to reduce sliding, thereby reducing wear, by redesigning the blocker shape.

Thermal Fatigue

Thermal cycling of the die surfaces during conventional hot forging results in the second most common reason for rejecting dies, namely heat checking. Thermal cycling (thermal fatigue) results from the intermittent nature of forging production.

The major factors influencing heat checking are:

- Die surface temperatures
- Surface stresses and strains
- Damage accumulation in thermal fatigue
- Microstructural effects of fatigue

Die Surface Temperatures. Information on die temperatures is best obtained from direct measurements. Surface temperatures for dies used in a mechanical press have been found to reach about 600 °C (1110 °F) in forging of steel cylinders that were preheated to approximately 1175 °C (2150 °F) and upset to 75% reduction in height (Fig. 24). Similar measurements have been made by Kellow, *et al.* (Ref 64), who upset medium-carbon steel samples in a slow hydraulic press and a HERF machine. Surface thermo-couples were placed at various distances from the axis of the 25 mm (1 in.) diameter billets. Experimental results showed that the temperatures obtained along the initial contact area of the workpiece and the die do not differ significantly between low- and high-speed forging. However, the temperatures obtained outside the initial contact area, where the billet surface extends during deformation, were significantly higher in high-speed forging (900 °C, or 1650 °F) as compared with low-speed hydraulic press forging (550 °C, or 1020 °F). These results are mainly due to differences in heat generation due to friction, which serves as one means of dissipating the energy produced by the forging machine.

Other measurements of die temperatures away from the die surfaces themselves demonstrate that large temperature gradients, as well as high temperatures, are induced in forging dies. These measurements include those of Voss (Ref 65), who measured temperatures in low-alloy (6F3) and chromium hot-work steel (H10, H12) radial forging dies preheated to 100 °C (210 °F) before forging (Fig. 25). Measurements away from the surface (at 0.5 mm, or 0.02 in., from the surface) show large temperature gradients. By comparing die temperatures during forging to those between forging blows (Fig. 25), it is apparent that very large temperature changes at the surfaces of forging dies may be expected as well. For this reason, large stresses and large strains due to temperature effects are experienced by the surface layers of forging dies.

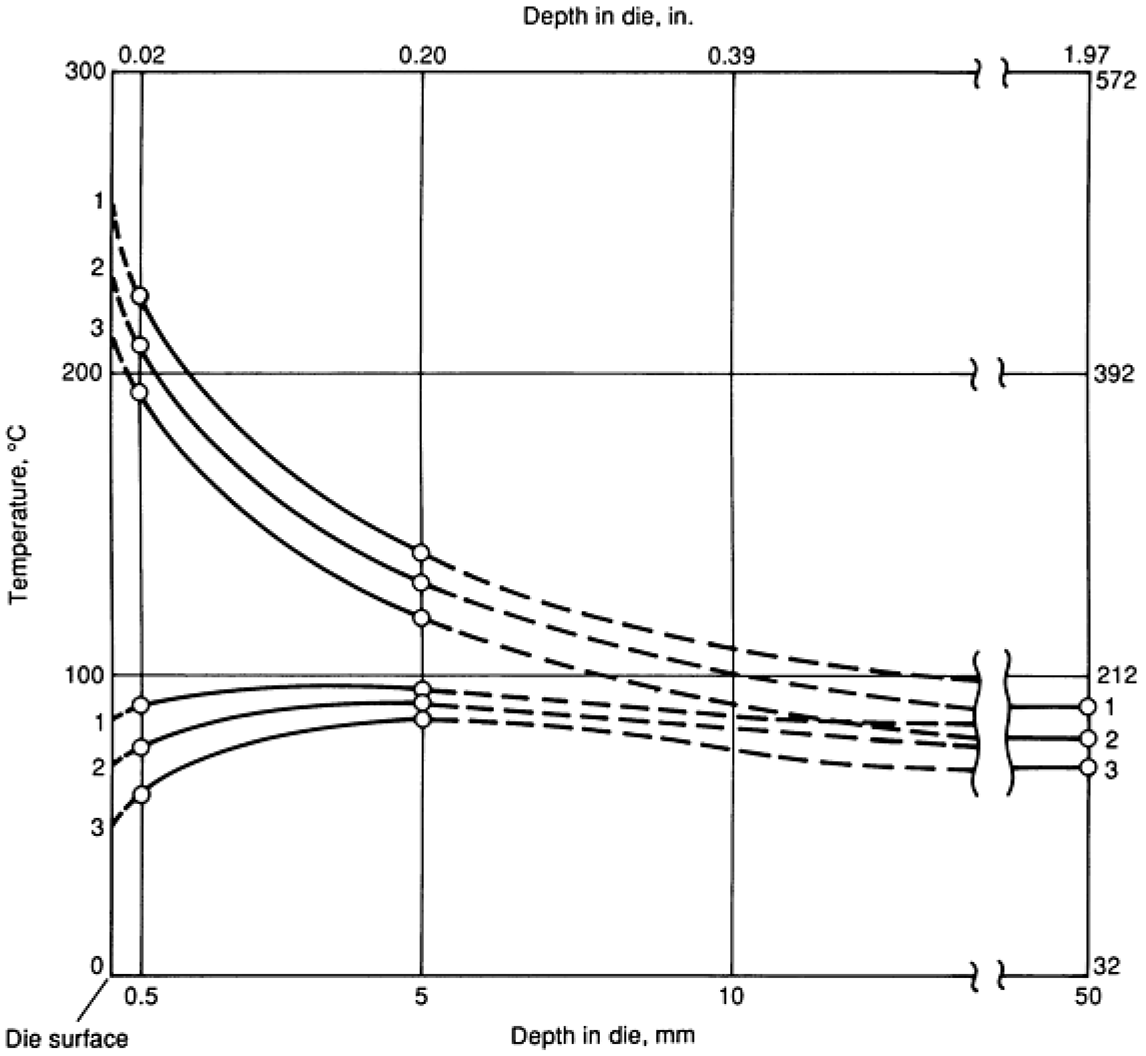


Fig. 25 Temperatures in dies with air-water cooling of the dies between blows. Initial die temperature: 100 °C (210 °F). Initial stock temperatures: (1) 1150 °C (2100 °F), (2) 1050 °C (1920 °F), (3) 950 °C (1740 °F). Upper curves are the temperatures achieved during forging; lower curves are the temperatures reached between forging blows. Source: Ref 65

Materials with high conductivity are less likely to develop large thermal gradients and fail by thermal fatigue than those with poor thermal conductivity. Although conductivity data for the various die materials are scarce, available measurements do show, for instance, that the tungsten hot-work die steels with higher conductivities should be more resistant to heat checking than the chromium hot-work die steels.

Surface Stresses and Strains. The stresses and strains that result from the temperature cycles experienced by the forging dies have two main sources: (1) thermal expansion and contraction, and (2) phase changes brought about by temperature cycling. The first of these is probably the easiest to quantify. This is because the thermal stresses and strains are approximately proportional to the maximum temperature difference (T_{max} - T_{min}) experienced by the dies and the thermal expansion coefficient of the die material. Most die steels have similar thermal expansion coefficients. Therefore, the thermally induced deformation of the dies is controlled primarily by the magnitude of T_{max} - T_{min}.

As might be expected, the tendency to heat check can be decreased by reducing T_{max} - T_{min}. This can be done in two ways. First, T_{min}, or the bulk die temperature, can be increased. However, such a change may adversely affect resistance to other forms of die failure. Alternatively, T_{max} can be decreased. The easiest way to do this is by decreasing the workpiece temperature or by using a lubricant with better thermal insulating properties.

Figure 26 shows the effects of increasing T_{min} or decreasing T_{max} on the fatigue life (in terms of number of cycles to produce a crack of certain length) of mild steel. It is seen that a 100 °C (180 °F) decrease in T_{max} is much more beneficial in extending the fatigue life than a similar increase in T_{min}. This result is generally true for die steels as well, and can be attributed to the greater reduction of the strain amplitude by decreasing T_{max}.

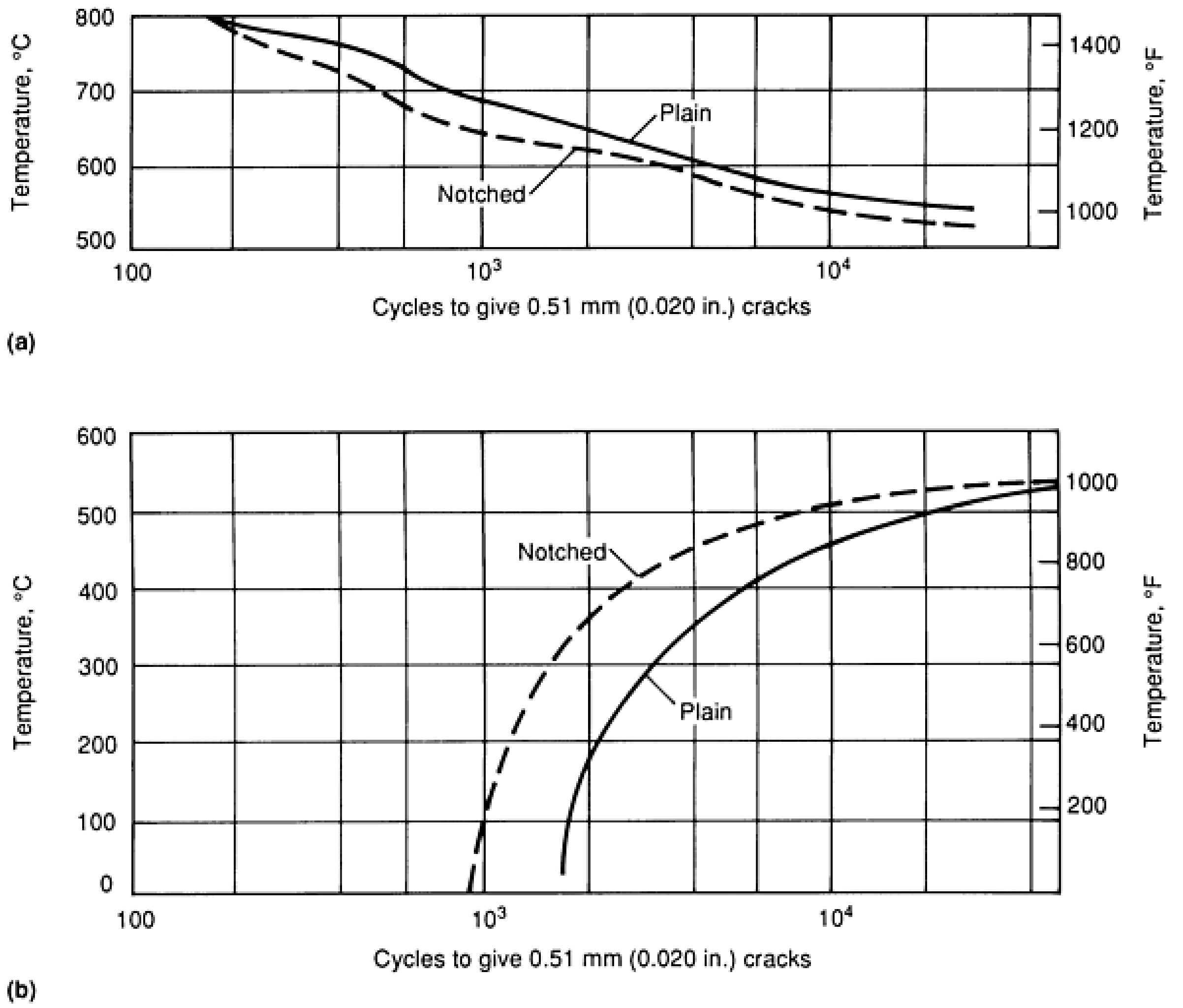


Fig. 26 Effects of maximum (a) and minimum (b) temperatures on the fatigue life of En 25 mild steel. Source: Ref 40

The effects of phase changes on thermal fatigue of forging dies has been examined by Rooks, Singh, and Tobias (Ref 66) and Okell and Wolstencroft (Ref 67). Both sets of investigators have concluded that die surface heating and cooling may lead to reversion of the tempered martensite to austenite and subsequent transformation back to martensite. Because austenite and martensite have different densities, such phase changes lead to strains and stresses that are imposed by subsurface layers of the dies that do not undergo the transformation.

As with the thermally induced strains, transformation-induced strains can be reduce either by keeping the maximum die surface temperature below the Ac_1 temperature (the temperature at which austenite forms, which is ~800 °C, or 1470 °F), or by keeping the minimum die surface temperature above the martensite start, M_s, temperature, which depends greatly on alloy composition, typical values being ~280 °C (~535 °F) for H11 and ~380 °C (~715 °F) for H21. Okell and Wolstencroft (Ref 67) suggested the latter possibility, but specified that it should only be used for the more highly alloyed die steels, which have good hot hardness because they resist overtempering.

Microstructural Effects on Thermal Fatigue. Because ductility has a large effect on the number of thermal cycles a forging die can undergo prior to forming cracks, microstructure can have a significant impact on the frequency of heat checking. The most important microstructural variables are cleanliness, grain size, and microstructural uniformity. Die

steels that are clean resist crack initiation inasmuch as inclusions act as nuclei for crack initiation. Thus, the use of a slightly more expensive steel that has been refined to remove inclusions may be a wise investment. Grain size can also affect thermal fatigue resistance because grain size has a large influence on crack initiation, with fine-grained material tending to perform better in this respect. Lastly, steels whose chemistry and microstructure are uniform (that is, free of segregation) tend to have uniform thermal properties (thermal expansion coefficients) and thus are able to resist thermal stresses and strains that may be developed due to such variations in a uniform temperature field.

Methods of Improving Resistance to Thermal Fatigue. The resistance to thermal fatigue can be improved by materials selection, lowering the maximum die temperature variations, or surface treatments.

Use of a Steel with Higher Yield Strength. Because thermal fatigue crack growth is controlled by the amplitude of the plastic strain increment, die steels with higher yield strengths, and thus higher elastic limits, are more resistant to thermal fatigue under a given set of process conditions.

Lowering the Maximum Die Temperature. Because thermal stresses and strains are related to the temperature changes that the die surfaces experience, decreasing the maximum temperature to which the die surface is exposed is beneficial. This can be accomplished by lowering the workpiece temperature or by using lubricants, such as glasses, that act as thermal insulators. Lowering the maximum die surface temperature is also helpful in avoiding transformation-induced strains, which, in conjunction with thermal strains, may cause thermal fatigue problems.

Raising the Bulk Die Temperature. Thermal stresses and strains and the tendency for heat checking can also be reduced by preheating the dies to higher temperatures. Use of this technique should be limited to die steels with good retention of hot hardness, such as the molybdenum hot-work die steels (H41 to H43).

Use of High-Quality Die Steel. Die steels that are clean, of fine grain size, and homogeneous in microstructure resist the initiation of fatigue cracks that are thermally or mechanically induced.

Use of Special Surface Finishes or Treatments. By eliminating machining marks, which act as stress concentrators, fatigue crack initiation can sometimes be avoided or delayed. Surface treatments such as nitriding or shot peening may also reduce thermal (and mechanical) fatigue problems by inducing residual compressive stresses into the surfaces of forging dies.

Mechanical Fatigue

Unlike thermal fatigue, the literature on mechanical fatigue of die steels is sparse. Perhaps the largest amount of data on this failure mechanism has been gathered by Thomas in a series of three-point bending experiments (Ref 68). Variables that he investigated included imposed load (of greatest importance in controlling fatigue behavior), hardness, material, position in the die block, strain rate, and temperature.

Methods of Improving Resistance to Mechanical Fatigue. Methods of minimizing failure due to mechanical fatigue fall into one of two categories. The first of these relates to die design and loading. In this area, redesign of dies (flash design) or performs to lower die stresses may totally eliminate problems of mechanical fatigue. Also, because of the logarithmic nature of fatigue-failure behavior, often only a slight decrease in applied loading can result in markedly improved fatigue lives. Such a reduction in load may be obtained, for example, through better control of the forging energy in hammers and slight modification of the flash configuration in die design.

The second major category of methods to improve fatigue resistance comes under the heading of material modification. Modifications include treatments (such as shot peening) that put the surface layers of the dies into compression. Another alternative is a total change of die material to one with a higher fracture toughness. Such a material can support larger fatigue cracks before total fracture occurs.

Catastrophic Die Failure/Plastic Deformation

The last two forms of die failure, catastrophic die failure and failure due to plastic deformation, will be discussed only briefly here. The first of these, catastrophic die failure, can be considered a special case of mechanical fatigue failure in which the fatigue life is only one cycle. It is usually a result of excessive forging stresses or improper die material/heat treatment selection or improper assembly of the die insert in the die holder. Forging stresses may be high because of excessive forging energy or because of the general shape of the forging die cavity (Ref 69). With regard to the former,

once the dies have come together in a forging operation, excessive forging energy can only be dissipated by elastic deformation of the dies themselves. Therefore, the fracture stress may be reached at points of stress concentration. Among the details of the forging die cavity that affect stresses are the draft angles and corner radii.

Improper die material selection can also result in catastrophic die failure. Dies for hard-to-work materials or in which there are points with high stress concentrations require die steels with good fracture toughness. These steels include low-alloy steels (such as 6F3 and 6F7) and some of the chromium hot-work die steels (such as H11). When the more highly alloyed die steels (such as H19 or H21) are employed (because of their wear resistance, for example), they should be tempered to lower-than-normal hardness in order to increase toughness if the dies are susceptible to catastrophic die failure.

Failure of forging dies to perform properly because of plastic deformation can be measured by hot hardness or yield strength. In general, the yield strengths of steels decrease with increasing temperature. However, yield strength is also dependent on the prior heat treatment, composition, and hardness. The higher the initial hardness, the greater the yield strength at various temperatures. In addition, the yield strengths of different die steels increase with alloy content--the tungsten hot-work die steels are harder than the chromium hot-work die steels, which are themselves harder than the low-alloy steels (Fig. 2 and 3). Not shown in Fig. 2 and 3 is the fact that the molybdenum hot-work die steels are even harder than the tungsten ones and thus manifest the greatest resistance to plastic deformation as far as die steels are concerned.

Surface Treatments and Coatings

A variety of die coatings and surface treatments are available to extend the lives of dies limited by wear. Among the most common coatings are plated chromium and cobalt-alloy deposits, which adhere to the die surface via a mechanical bond, and weld fusion deposits, which entail a metallurgical bond. The latter can be used in rebuilding excessively worn dies as well. Recently, the use of ceramic coatings (for example, carbide and nitride coatings) applied by processes such as chemical vapor deposition (CVD) have been found to extend forming die lives.

The surface layers of ferrous forming dies can also be hardened by alloying them with nitrogen or boron. Nitriding is the most common of the surface treatment techniques and can be accomplished using a gaseous, liquid, or plasma medium. The plasma technique (ion nitriding) appears very attractive because the formation of brittle white layers, which are unavoidable in other nitriding processes, can be eliminated or at least minimized. Boriding of die surfaces can also be accomplished using a variety of media, and increases in die life of the same magnitude as those obtained with nitriding have been reported.

Coatings have been used extensively in net shape forming to reduce friction and wear. Coatings can be applied to either the workpiece or the die. In the case of the workpiece, the coatings are made of soft material with good adhesion, lubricity, and low shear strength. In cold forming, for example, phosphate coatings are used to reduce interface friction and die wear (Ref 70). Resin-bonded coatings containing solid lubricants have also been successfully used (Ref 71, 72, 73). In hot-forming applications, hard coatings are generally used. They are applied to the die or mold surface by mechanical, thermal, or chemical means.

A hard surface layer reduces the frictional force and the wear rate when sliding against a relatively soft workpiece material if the coating/workpiece material pair is chemically stable and the coating is well bonded and mechanically compatible with the substrate (die material) (Ref 74). The role of the hard layer is to prevent plowing, whereas chemical insolubility is needed to ensure minimal dissolution. Hard coatings are especially useful when the dominant wear mechanism is abrasive wear.

Typical hard coating materials are oxides, carbides, nitrides, borides, and amorphous glasses. One of the main considerations in the choice of a coating material is the quality of bonding between the coating and die material (substrate). The bonding could be chemical, mechanical, or both. Chemical bonding is caused by a reaction or diffusion of atoms between the coating and the substrate to form a solid solution at the interface.

Alloying Surface Treatments

The lives of steel forming dies limited by wear can often be increased by various surface treatments in which the structure of the surface is alloyed with, for example, nitrogen, carbon, or boron. Other common methods of surface hardening of steel (for example, flame hardening and induction hardening) have not been reported in the literature as having been used

for forming dies. This is perhaps due to the large loss in toughness and distortion of the die cavity that these methods may cause.

Nitriding is probably the most common treatment for hardening the surface layers of forging dies. It is useful for applications in which the surface temperature does not exceed 565 to 595 °C (1050 to 1100 °F) in service. As with most of the surface treatment processes, nitriding finds its greatest application for press-forming dies in which strength and wear resistance are more important than toughness. However, there are reports of nitriding being successful in impact applications such as hammer forging (Ref 75). Wear rates have been reduced by as much as 50% using nitriding (Fig. 27) (Ref 75, 76).

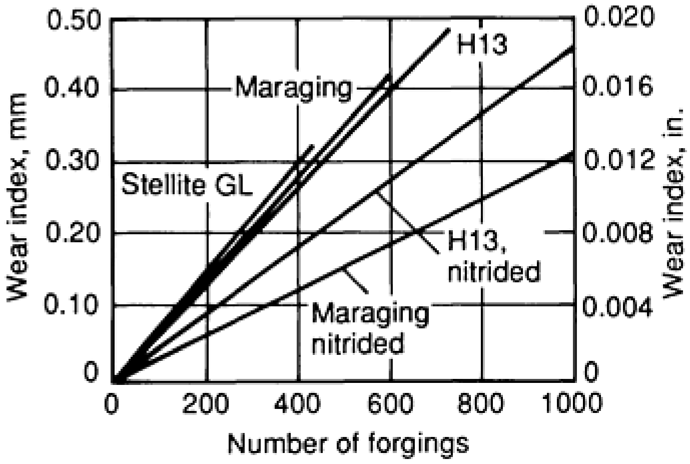


Fig. 27 Relative wear rates of nitrided ad non-nitrided tool steels used in extrusion forging Source: Ref 75

Nitriding processes are performed at temperatures between 495 and 565 °C (925 and 1050 °F). It is important that tempering of the die steel be performed at a temperature exceeding the nitriding temperature prior to nitriding in order to optimize the property combination of the core and the surface of the dies. Also, because of the low nitriding temperatures, there is generally little distortion from this heat-treating process.

Although the depth and hardness of the nitride case depends a great deal on the nitriding time, these properties (particularly the hardness) are sharply dependent on the composition of the steel as well. Die steels containing large amounts of strong nitride formers such as chromium, vanadium, and molybdenum form shallow, very hard surface layers. On the other hand, low-alloy chromium-containing die steels (such as 6G and 6F2) form deeper surface layers that are tougher, but not as hard.

Detailed information on specific techniques for nitriding--gas nitriding, liquid (salt bath) nitriding, and ion nitriding--can be found in Volume 4, *Heat Treating,* of the *ASM Handbook.*

Boriding and Carburizing. Boron can be added to surface layers by a diffusion treatment that can be carried out in either gas, molten salt, or pack media at a temperature between 900 to 1100 °C (1650 to 2010 °F), depending upon the process and the material to be borided. Extremely hard surface layers with low coefficients of friction are formed, provided the base metal forms borides. The process does not require quenching. If the base material has to be heat treated, the heat treatment can be done after boriding, although care is required to reduce quenching stresses to prevent spalling of the borided layer.

Very little is known about the usefulness of boriding forming dies. Vincze (Ref 77) claims a 70% increase in die life with dies surface treated by boriding compared to untreated dies. In this study, boriding was carried out by filling the die impression with a mixture of 10% B_4C, 40% sodium borate, and 50% hardening salt, and heating the pack at 900 °C (1650 °F) for 3 h. This diffusion heat treatment was followed by quenching and tempering. Burgreev and Dobnar (Ref 78) also report large increases in hammer forging die life when boriding is used.

Boriding of steels is also done electrolytically. Boron atoms are electrodeposited onto the metal from a bath of molten salt containing fluorides of lithium, sodium, potassium, and boron. The dies are borided in the 800 to 900 °C (1470 to 1650 °F) temperature range in an atmosphere of argon or a mixture of nitrogen and hydrogen. Thickness of coating is from 0.013 to 0.05 mm (0.0005 to 0.002 in.), and treatment lasts 15 min to 5 h (Ref 79).

It has been stated that boriding results in undesirable interaction with alloying elements of hot-work die steels (H series) and develops a soft layer (Ref 78). Porosity in the borided layer can develop for steels that require postboriding heat treatment. For this reason, it is preferable to limit boriding to those alloys that do not require further high-temperature treatment. For example, A6 air-hardening steel can be hardened from the boriding temperature by cooling in air, and only requires tempering. This steel, therefore, can be safely borided.

Carburizing of hot-work die steels is uncommon and not popular for two main reasons (Ref 80). As with boriding, the high temperatures required for carburizing (815 to 1095 °C, or 1500 to 2000 °F) lead to distortion of the dies on cooling. Secondly, the high-carbon surface layer, although it greatly increases hardness, can drastically reduce the toughness of the dies.

Additional information on boriding and carburizing can be found in Volume 4, *Heat Treating,* of the *ASM Handbook.*

Ion implantation is an alloying surface treatment that has been recently applied to forming dies. The process was first developed in the late 1960s to introduce electrically active elements into semiconductors of microelectronic devices. Now, more than 2000 commercial ion implantation systems are being used worldwide for semi-conductor processing (Ref 81).

Ion implantation research has increased steadily since the first publication of Hartley, *et al.* (Ref 82). The description of the process, its advantages and shortcomings, and the industrial application of the process have appeared in several review articles (Ref 83, 84, 85, 86, 87, 88, 89). Some ion implantation applications in forming are discussed below.

Ion Implantation of Stamping and Cutting Tools. The number of applications involving the reduction of wear in stamping and punching tools probably exceeds all other uses of ion implantation for metals. The dominant cause of failure in high-speed tool steels used in metal stamping is adhesive wear. Ion implantation of titanium and carbon has provided the optimum treatment for such tools. For example, the life of punches and dies for manufacturing aluminum beverage cans has been improved by 6 to 10 times as compared with that of untreated tooling. Similarly, punches for pressing powders into pellets have lasted twice as long after being treated. These applications range from forming pharmaceutical pills to more demanding service such as pressing ceramic pellets. Selected examples are listed in Table 9.

Table 9 Examples of extending tool life via ion implantation

Tool	Material	Application	Results
Chasers	Tool steel	Cutting threads in 380 aluminum	5× life
Draw die	D2	Drawing 3.5 mm (0.140 in.) hot-rolled steel	22× life before polishing
Flanging ring	D2	2 mm (0.080 in.) hot-rolled steel (flywheels)	80× life before polishing
Forming dies	W1	1 mm (0.042 in.) cold-rolled steel	Pickup reduced, improved finish on part
Forming	D2	Forming electrical connectors	At least 60% longer life
Pierce punches	M2	13 mm ($\frac{1}{2}$ in.) hot-rolled steel	12× life
	M2	1.3 mm (0.050 in.) Hastelloy (jet engine part)	30% longer life
Pilot pins	M2	0.08 mm (0.003 in.) 301 stainless steel	5× life
Plastic forming	Stainless steel	Shaping vinyl siding	Wear rate reduced for dies
Plug gage	O6	Gaging powdered metal part	At least 2× life
Reamers	Carbide	Finish reaming gray cast iron	2× life
Slitters	Carbide	Cutting rubber (belts)	Friction reduced 30%
Sprue bushings	P20 (chromium plated)	Injection molding thermoset (20% glass)	At least 4.5× life
Taps	M2 (chromium plated	Tapping cold-rolled steel nuts	8× life
Thread form taps	M2	Tapping 380 aluminum	2× life
Ultrasonic electrodes	D2	Welding aluminum to steel	2.5× life (mechanical wear)
Wire compacting	Inconel, M2	Compacting copper-base composite wire	At least 16× life

dies

Source: Ref 90

Ion implantation has also been used on a variety of low-temperature (to 200 °C, or 390 °F) cutting and slitting tools. Knife blades for cutting synthetic fibers have been implanted with titanium and carbon to improve wear resistance. Other applications include slitting blades for cutting rubber, hay baling cord cutters, plastic bandage cutter tools, corn-husking blades, paper slitters, and paper punches. Typical life improvements of 6 to 10 times have been reported for most of these applications.

Ion Implantation of Injection Molds. Another of the better-known application areas for the ion implantation process is that of injection molds for plastics and metal powders. Simple implantation with nitrogen has provided several-fold improvements in the life of these tool steel molds, along with improved mold-release characteristics and reduced corrosion from chemical additives. Molds used for powder metallurgy have also shown less tendency to fracture than untreated molds.

Ion Implantation of Extrusion and Wire-Drawing Tools. Ion implantation has proved effective in treating components such as spinnerettes for nylon or rayon fibers, which have numerous tiny orifices. The usual failure mechanism in these tools is a combination of corrosion and abrasive wear. Implantation of titanium and carbon into the spinnerettes (usually made from stainless steel) has improved the resistance to the abrasive particles (titanium oxide) that are added to the synthetic material. The treated spinnerettes have also been more resistant to the corrosive conditions encountered during the cleaning process. Ion implantation of wire-drawing dies (up to 25 mm, or 1 in., in diameter) has not only increased the life of the dies, but also improved the surface finish of the drawn wire.

References 90, 91, 92, 93, 94, 95 deal specifically with application of ion implantation in forming. Additional information can also be found in the article "Ion Implantation" in this Volume.

Plasma source ion implantation (PSII) is an ion implantation technique recently developed by J.R. Conrad of the University of Wisconsin--Madison (Ref 96). Objects to be implanted are placed directly in a plasma source and then pulse-biased to a high negative potential. A plasma sheath forms around the target and the ions bombard the entire target simultaneously (Ref 97, 98, 99).

Carbide Coating by Toyota Diffusion Process. Good surface covering and strongly bonding carbide coatings, such as VC, NbC, and Cr_7C_3, can be formed on die steel surfaces by a coating method developed at Toyota Central Research and Development Laboratory, Inc. of Japan (Ref 34).

In the Toyota Diffusion (TD) process, metal dies to be treated are degreased, immersed in a carbide salt bath for a specific time period, quenched for core hardening, tempered, and washed in hot water for the removal of any residual salt. The borax salt bath contains compounds (usually ferroalloys) with carbide-forming elements such as vanadium, niobium, and chromium. The bath temperature is selected to conform to the hardening temperature of the die steel. For example, the borax bath temperature would be between 1000 and 1050 °C (1830 and 1920 °F) for H13 die steel.

The carbide layer is formed on the die surface through a chemical reaction between carbide-forming elements dissolved in the fused borax and carbon in the substrate. The carbide layer thickens due to reaction between the carbide-forming element atoms in the salt bath and the carbon atoms diffusing into the outside surface layer from the interior of the substrate.

The thickness of the carbide layer is varied by controlling the bath temperature and immersion time. An immersion time of 4 to 8 h is needed for H13 steel to produce carbide layers with satisfactory thickness (5 to 10 μm) for die-casting applications. Dies are then removed from the bath and cooled in oil and salt or air for core hardening followed by tempering.

The salt bath furnace consists of a steel pot with heating elements; no protective atmosphere is needed. Selective area coating is accomplished by the use of copper or stainless steel masking, plating, thermal spraying, or wrapping with foils. The type of carbide coating can be changed easily by using a different bath mixture or more than one pot, each containing different carbide mixtures.

The process is applicable to most steels and some nonferrous metals. Satisfactory results have been obtained for H12 and H13 steels, which are the most widely used steels for die-casting dies. Coated steels exhibit high hardness and excellent resistance to wear, seizure, corrosion, and oxidation. In addition, resistance to cracking, flaking, and heat checking is claimed. Hardness of the coating depends on layer composition; 3500 HV for vanadium carbide, 2800 HV for niobium carbide, and 1700 HV for chromium carbide.

Additional information on the TD process can be found in Ref 100 and in Volume 4, *Heat Treating,* of the *ASM Handbook.*

MetalLife Treatment

A new surface treatment for dies and molds, MetalLife, utilizes a controlled micropeening method to develop rough texture and beneficial compressive stresses on the surface of dies and molds (Ref 101, 102, 103, 104). Rough texture increases die-lubricant retention characteristics and the residual compressive stresses inhibit the initiation of fatigue microcracks due to the mechanical or thermal cycling of die surfaces during operation.

Shotpeening of surfaces is not a new concept, but the conventional process is not very suitable for die and molds because it results in pitting and stress raisers on the surface that can result in premature tooling failure. Uncontrolled use of a smaller blasting media can result in erosive and abrasive wear. The key variables that need control are the media size, concentricity, angle of impingement, velocity, and dwell time for each media used. Sometimes multiple treatments have to be performed.

Punch and die life increases of 6 to 10 times are claimed using the MetalLife treatment. Drawing and forming dies have shown life improvements of 10 to 20 times in certain cases. In the field of die casting, MetalLife has been used to close heat-checking cracks and increase the life of dies.

Ceramic Coatings

There are a number of ceramic coatings that can be applied by various means to metal parts to improve their service properties; these ceramic materials are electrically nonconductive, have up to 20 times the abrasion resistance of metals, and can withstand temperatures in excess of 2480 °C (4500 °F) (Ref 105). Among the many ceramic wear-resistant materials available for coatings are titanium carbide, titanium nitride, and chromium carbide. These materials can be applied to chromium hot-work steels and the air-hardening tool steels by the chemical vapor deposition (CVD) process (Ref 106). In this process, the metal part to be coated is placed in a special reactor vessel after which it is heated and reacted with gas containing coating materials species. Selection of suitable coatings and metals depends strongly on the compatibility of the two from a thermal expansion viewpoint. If the expansion coefficients are widely different, the coating may crack when the part is cooled to room temperature. Because a surface interdiffusion layer is also produced, the possibility of forming soft or brittle compounds must also be considered. From previous discussions, it is known that these considerations are also important from the perspective of expected performance during forging. A good match of thermal properties is required to prevent heat checking, and tough, hard surface layers are needed to offer resistance to wear and brittle fracture. Besides being hard, these coatings generally have good lubricity. However, because of oxidation problems, they must be used at temperatures below 650 °C (1200 °F), which is above the typical operating temperature of forming dies made of hot-work die steels. To date, CVD coatings have been used for forging dies to a limited extent only. In one application, use of a TiN coating on H26 press-forging dies increased die life from approximately 18,000 (uncoated) to 51,000 pieces, as compared to a die life of 36,000 pieces using chromium-plated dies (Ref 107). The use of these coatings is sure to increase in the future, particularly in applications where life is limited by abrasive wear.

Evidence that ceramic surface coatings can extend the life of hot-forging dies has also been demonstrated in die wear trials on H12 (Ref 108, 109). In these trials, die life increases in excess of 10% were reported.

Other ceramic coatings that may find hot-forming application include alumina, zirconia, chromium oxide, and magnesium zirconate. All of these coatings have excellent wear resistance, especially chromium oxide, which has a diamond pyramid hardness of 1200 kg/mm^2 with a 300 g load; aluminum oxide has a slightly lower hardness, 1100 kg/mm^2 with a 300 g load (Ref 105). Several ceramic coatings provide excellent thermal resistance in excess of 2480 °C (4500 °F). Zirconia has a melting point close to 2480 °C (4500 °F) and is extremely resistant to thermal shock. Magnesium zirconate, with a melting point near 2150 °C (3900 °F), and yttria-stabilized zirconia, with a melting point near 2650 °C (4800 °F), are used as thermal barrier coatings.

Ceramic coating application requires the following steps:

- Cleaning
- Roughening
- Undercoating
- Coating
- Finishing

Surface preparation (cleaning, roughening, and undercoating) is the most critical step and determines application success. Mechanical, metallurgical-chemical, and physical bondings play an important role in coating adherence and strength.

Coatings are applied on the substrate by four common methods (in addition to the CVD method described above):

- Oxygen acetylene powder
- Oxygen acetylene rod (welding)
- Plasma spraying (torch)
- Detonation gun

In the oxygen acetylene powder method, the powder is fed into a flame at 2760 °C (5000 °F) and sprayed on the substrate by compressed gas. The coatings produced by this method are generally porous with low adhesion. This process is of moderate costs.

In the oxygen acetylene rod method, fused ceramic material in a rod form is introduced into a 260 °C (500 °F) oxyacetylene torch. Molten ceramic is sprayed at speeds up to 170 m/s (550 ft/s) via compressed gas on the target. This results in a coating with high cohesive bonding.

In plasma spraying, ceramic powder is introduced into a plasma (ionized gas) at temperatures as high as 16,650 °C (30,000 °F). The high-pressure plasma gas accelerates molten particles on the target. This method produces well-bonded high-density coatings, but is very expensive.

Extremely dense coatings are produced by the detonation gun process. This process is preferred when tungsten carbide coatings are to be applied. An explosion of oxygen and acetylene gases produces 3315 °C (6000 °F) temperatures, melting the ceramic and producing a molten jet that impinges the target at speeds up to 760 m/s (2500 ft/s).

A novel method for coating metalworking dies with refractory metals has been patented by a group of researchers at United Technology Corporation (Ref 110). In this method, a refractory metal coating is sprayed by a plasma gun and subsequently compacted under conditions of minimum shear stress. Refractory metals selected include molybdenum, niobium, tantalum, tungsten, rhenium, and hafnium because they have melting points in excess of 2200 °C (4000 °F) and sufficient ductility for compaction. The compaction of the coating is achieved by processing a workpiece, which has previously been formed to the end shape, through the die. This pressing of a preformed workpiece reduces the metal flow and shear stresses to a minimum, thereby avoiding shear and spall of the coating. Refractory-coated H13 tool steel dies have exhibited significant improvements in wear resistance.

Electroplating

Chromium Plating. Chromium is usually applied to metal pieces using electroplating baths composed of chromic acid and some sulfate or fluoride compound (Ref 111). Bath temperatures are between 45 and 65 °C (110 and 145 °F). The kind of plating used for forging dies is called hard chromium plating and results in surface deposits typically between 25 and 500 μm (1 and 20 mils) thick. After application of the plating, the die cavity is ground to finish dimensions.

Chromium plating has been used to a modest extent on industrial forging dies, but there is conflicting evidence as to its value. Dies with deep cavities, sharp corners, or projections that show cracking due to thermal or mechanical fatigue should not be chromium plated. This may be due to the tendency of hard chromium platings to contain microcracks that

open during cyclic loading. On the other hand, dies for thin forgings that must be discarded because of wear (especially at the flash land), and best suited for chromium plating (Ref 112)

Cobalt Plating. As with chromium, various cobalt alloys have been applied to hot-forging die steels, primarily to extend life through reduction of die wear (Ref 113, 114, 115, 116, 117). Typically composed of alloys of cobalt and tungsten or cobalt and molybdenum, these coatings are applied in electroplating baths using so-called electroplating brushes, which allow a small or selected area to be plated. Besides offering improvements in wear resistance, it appears that these coatings also possess a low coefficient of friction and can in some cases solve problems involving sticking of the workpiece. Many of the criticisms often leveled against coatings, namely loss of adhesion or flaking due to poor resistance to shock loading, appear not to apply to these coatings. Hence, it is not surprising that increases of die life up to 100% in press-forging operations are not uncommon with the use of these coatings. The kinds of parts for which these improvements have been obtained are varied and include gear levers, turbine blades, and suspension end-sockets for cars.

Hardfacing

Hardfacing is a weld fusion process that produces deposits which are metallurgically bonded to the substrate. In the early days, hardfacing was used for repair and maintenance of die and molds. It is now being used increasingly as an inexpensive means for depositing a hard layer on localized wear-prone die areas.

For dies and molds, these deposits have the following applications:

- Deposits of identical material onto a die block to repair it or to allow resinking of it
- Deposits of higher-alloy steels (for example, chromium hot-work steels) onto the die surface of low-alloy steels to improve the service performance of the dies
- Deposits of hard or high-temperature materials (usually cobalt- or nickel-base alloys) onto low-alloy or hot-work steels to improve the service performance of the dies. These alloys come under the general heading of hardfacing or hard-surfacing alloys

Hardfacing Processes. Before discussing specific alloys, the processes by which they are deposited will be briefly reviewed. The first step in any of the hardfacing processes should be the annealing of the die block into which the rough impression has been sunk (Ref 118). This relieves residual stresses and helps prevent cracking during welding of the surface layer. After annealing, the die block should then be reheated to a temperature of 325 to 650 °C (600 to 1200 °F), which is also necessary to minimize cracking due to thermal gradients set up between the surface and the interior during welding. The application of the surface layer can then be performed by one of a number of welding processes (Ref 119, 120):

- Gas torch welding (combustible gas welding)
- Manual arc welding
- Submerged arc welding
- Gas shielded arc welding (TIG or MIG)
- Open arc welding
- Thermal spraying
- Fusion treatment
- Plasma spraying (plasma arc welding)
- Transferred arc plasma
- Flame plating
- Deposition process (electroslag welding)

Together with the solidification conditions, the amount of melted base material and base-material dilution is important for wear properties. Hardfacing methods differ considerably from each other and also compare with powder spray methods, which show almost no base-material dilution due to mixing.

Combustible-gas welding offers many advantages in depositing smooth, precise surfaces of high quality. This is done by using a carburizing flame that causes "sweating," or welding of a thin surface layer that spreads freely and prevents metal

buildup. For repair of dies, the shielded metal arc method is preferred. It allows high productivity and has the advantage of low heat input and thus minimal distortion of the die cavity.

After welding, the die block must be cooled to room temperature to prevent cracking of the weld deposit. The die impression is then finished, machined, and ground. Heat treatment (austenitizing, quenching, and tempering) of the die block is performed last. Once again, differences in thermal properties between the base metal and surface deposit are critical insofar as thermal cracking is concerned. Because this is also an important consideration in the performance of forming dies, it is not unusual that welding alloy suppliers are sometimes hesitant about recommending many combinations of die block material and hardfacing alloy (Ref 121).

Hardfacing Alloys. For hardfacing, welding alloys are generally based on iron, cobalt, or nickel. Hard phases are formed by addition of carbon (in iron) or boron (in nickel). The volume fraction of hard phase is very important for the wear resistance in the weld deposit. Often there is no proportional dependence and the best wear resistance is not achieved by the highest hard-phase concentration.

Various ferrous alloys are used to repair steel dies or to lay down deposits with better wear and heat resistance than the substrate. These alloys are very similar to the low-alloy and hot-work tool steels in composition. Austenitic and austenitic-ferritic materials are preferred for wear resistance under heavy loads.

The use of nickel- and cobalt-base alloys in hardfacing offers a considerable cost savings over die blocks made of these alloys. In a typical hardfacing operation, a one- or two-alloy layer, each about 0.25 to 1.25 mm (0.010 to 0.050 in.) thick, are deposited on the die. If a large amount of buildup is desired or required, however, it is advisable to apply layers of stainless steel or low-alloy filler metal first rather than many layers of the more expensive nickel or cobalt hardfacing materials.

Detailed information on the methods for depositing hardfacing alloys and their resistance to wear can be found in the article "Friction and Wear of Hardfacing Alloys" in this Volume.

Electro-spark deposition (ESD) is a variation of hard surfacing that has been used extensively in Europe for improving the galling resistance of material (Ref 122). Electrodes of WC, TiC, and Cr_3C_2 materials have been deposited on type 316 stainless steel and other substrates. The ESD process has been found to be effective in fusing metallurgically bonded coatings to the substrate at low heat with the substrate remaining near the ambient temperature.

Hard Coatings for Cold Extrusion

In the cold extrusion of steel, compressive loads up to 3000 MPa (435 ksi) and tensile loads up to 1500 MPa (220 ksi) are not unusual. Core hardening (or through hardening) of the tool steel makes the punches brittle, leading to early failures. The highly stressed tools, therefore, either should be made from tungsten carbide (a costly material) or they should be hard coated.

A very detailed study of chemical and physical treatments for backward can extrusion has been reported by Westheide, *et al.* (Ref 123). They divided the surface treatments used into reaction- and coating-layer processes (Fig. 28). In the reaction layers, the layer element diffuses into the substrate (die material); in the coating layers, the primary adhesion mechanism is mechanical interlocking. Figure 28 also identifies the type of coating that can be deposited on the substrate.

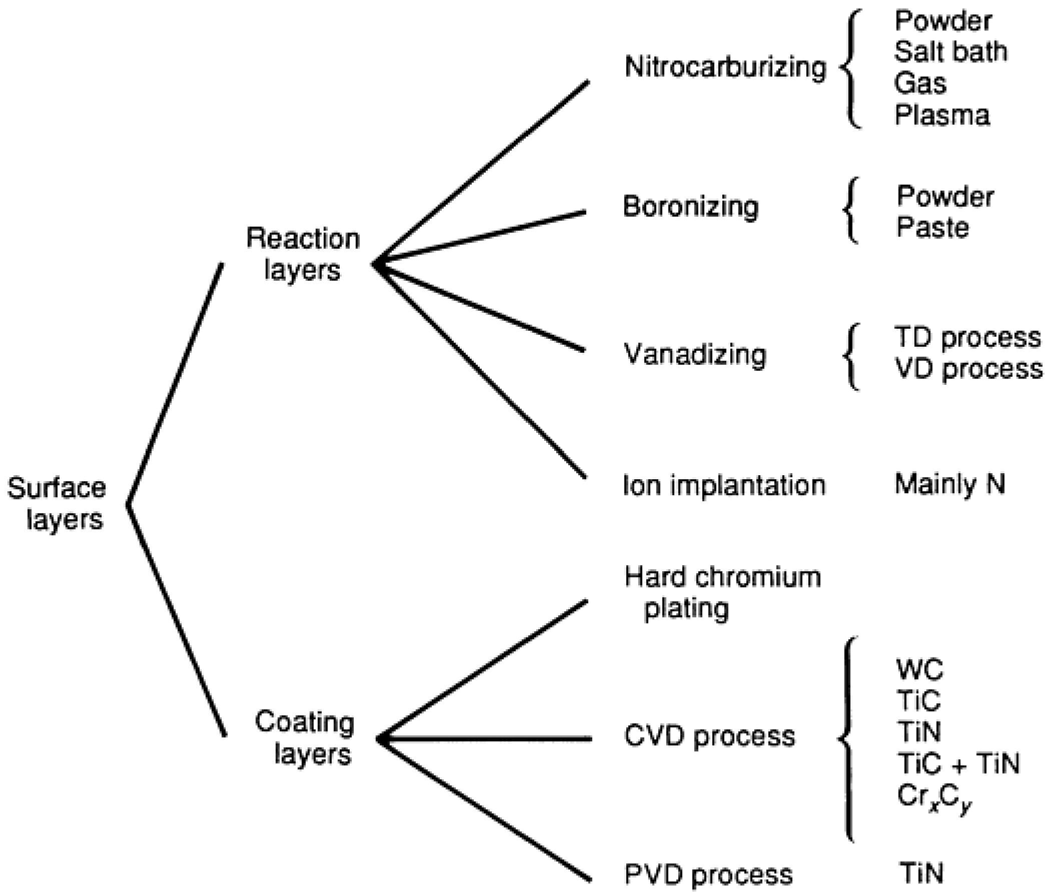


Fig. 28 Coating processes carried out on die materials used for cold extrusion. Source: Ref 123

In this study, billets of case hardening steels (similar to AISI 5120) were used. Punches were either made of a cold-working tool steel (similar to AISI D2) or of high-speed tool steel (similar to AISI M2); the former hardened to ~62 HRC and the latter to ~64 HRC. Backward can extrusion experiments were carried out on a 630 kN (70 ton) press at 40 strokes/min. The billet height to internal diameter ratio was unity and soap was used as the lubricant.

A comparison of treatments for backward can extrusion is given in Fig. 29 for 10,000 extrusions. Lowest wear rates were obtained for PVD TiN coatings and TD vanadium carbide (vanadized) coatings.

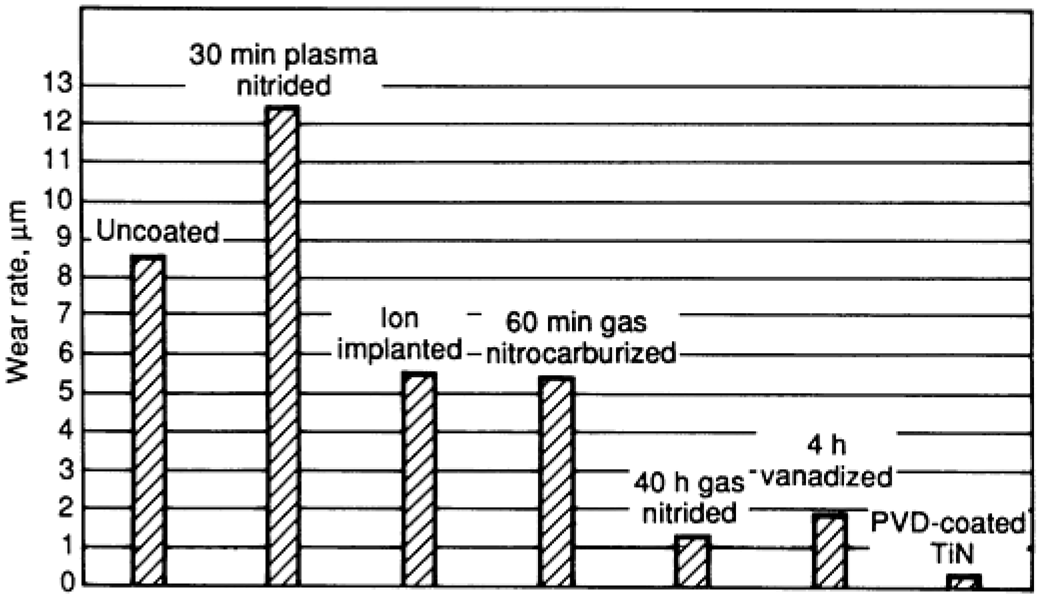


Fig. 29 Comparison of coatings for backward can extrusion. Source: Ref 123

The application of TiN coatings in cold extrusion is reviewed in Ref 123 and 124. Some of the cold extrusion applications where TiN coatings have proved beneficial are hexagonal socket press tools, tools for spur gear teeth, and hydraulic valve stem housings (Ref 123). Vanadium carbide coatings deposited by the TD process have been applied to extrusion dies in rubber forming. Die life increases from 30 h for hardened steels to 900 h for coated steels have been found in the production of rubber window seals (Ref 104).

Surface Treatments and Coatings for Cold Upsetting

Westheide (Ref 123) also carried out similar studies on protective wear-resistant coatings for upsetting dies. A qualitative comparison of various coatings is provided in Table 10. Hard chromium plating and TiC coatings via CVD were found to have excellent wear resistance.

Table 10 Comparison of selected coatings for cold upsetting

Coating	Surface after coating	Wear reduction	Layer adhesion	Remarks
Nitriding and nitrocarburizing	P	G	M	Chipped off
Vanadizing	M	M	G	Layer polished, uncertain measurement
Ion implantation	E	G	E	B^+ implantation = unsuccessful
Hard chromium plating	E	E	E	. . .
CVD-W_2C	P	G	P	Layer chipped off
CVD-TiC	P	E	E	. . .
PVD-Ti*X*	M	M	P	Layer chipped off due to interruption of coating

Note: E = excellent, G = good, M = moderate, and P = poor.

Source: Ref 123

References

1. V. Nagpal, Battelle Columbus Laboratories, unpublished research, 1976
2. J. Del Rio, "How To Select Die Steels for Forging," unpublished manuscript

3. V.A. Kortesoja, Ed., *Properties and Selection of Tool Materials,* ASM International, 1975
4. "Tool Steels," Universal-Cyclops Steel Corporation
5. "Hot-Work Die Steels," A. Finkl and Sons Company
6. "High Speed Tool and Die Steels," Latrobe Steel Company
7. "Survey of Hot Work Steel Grades," VEW (Bohler Brothers of America, Inc.)
8. "Working With AL Tech Tool Steels," AL Tech Specialty Steel Corporation
9. L.R. Cooper, Hot Work Die Steels for Closed-Impression Forging, *Precis. Met. Molding,* Vol 25 (No. 4), Apr 1967, p 46; Vol 25 (No. 5), May 1967, p 69; Vol 25 (No. 6), June 1967, p 74
10. G.A. Haynes, General Discussion on Maraging Steels, *Tools and Dies for Industry,* The Metals Society, London, 1977, p 415
11. T.A. Dean and C.E.N. Sturgess, Warm-Forming Practice, *J. Mech. Work. Technol.,* Vol 2, 1978, p 255
12. A. Torry, Casting by the Tru-Process Method, *Met. Treat.,* Vol 29 (No. 206), Nov 1962, p 451
13. G.D. Stutzman, Cast Tooling Gets Better, *Metalwork. Prod.,* Vol 113, 6 Aug 1969, p 29
14. G.D. Stutzman, "Cast Forging Dies," Paper CM69-236, American Society of Tool and Manufacturing Engineers, 1969
15. *Properties and Selection: Stainless Steels, Tool Materials and Special Purpose Metals,* Vol 3, 9th ed., *Metals Handbook,* ASM International, 1980
16. "High-Temperature, High-Strength, Nickel-Base Alloys," The International Nickel Company, Inc., 1977-1978
17. V. Nagpal and G.D. Lahoti, "Selection of Die and Mandrel Materials for Radial Forging," Interim Topical Report on Contract No. DAAA22-78-C-0109, Battelle Columbus Laboratories, Jan 1979
18. K.C. Antony, Stellite Division, Cabot Corporation, private communication, 1981
19. "Properties and Proven Uses of Kennametal Hard Carbide Alloys," Kennametal, Inc., 1978
20. R.L. Lewis, Y.L. Chen, C.C. Chen, M.-S. Han, and H.S. Kim, "State of the Technology in Die Casting," Report No. ERC/NSM-87-12, Engineering Research Center for Net Shape Manufacturing, July 1987
21. S. Malm and J. Tidlund, "Increased Life for Die Casting Dies," Paper No. G-T79-051, 10th SDCE International Die Casting Expo and Congress (St. Louis), Society of Die Casting Engineers, 1979
22. W. Young, "Why Die Castings Fail," Paper No. G-T79-092, 10th SDCE International Die Casting Expo and Congress (St. Louis), Society of Die Casting Engineers, 1979
23. J.F. Wallace, "Effect of Selected Processing on Compositional Factors on the Failure of H-13 Steel as Dies for Aluminum Die Casting," available from Case Western Reserve University Library
24. B. Johansson, L.-A. Norstrom, and N. Ohrberg, Thermal Fatigue and Thermal Shock Testing of Some Hot-Work Tool Steels for Die Casting, *Towards Improved Performance of Tool Materials,* Proceedings of International Conference, The Metals Society, London, Apr 1981
25. E. Krainer, A. Schindler, J. Kiefer, and G. Griggs, Wear and Creep Properties of Hot Work Steels for Extrusion and Die Casting Applications, *Towards Improved Performance of Tool Materials,* Proceedings of International Conference, The Metals Society, London, Apr 1981
26. H. Tabe, Thermal Shock Cracking of Tool Steel, *Tool Materials for Molds and Dies,* Proceedings of International Conference, 30 Sept-2 Oct 1987 (St. Charles, IL), G. Krauss and H. Wordberg, Ed.
27. J.C. Benedyk, D.J. Moracz, and J.F. Wallace, "Thermal Fatigue Behavior of Die Materials for Aluminum Die Casting," Paper No. 111, 6th International Die Casting Symposium and Congress, Society of Die Casting Engineers, 1970
28. L. Northcott and H.G. Baron, The Craze Cracking of Metals, *J. Iron Steel Inst.* (London), Vol 184, Part 4, Dec 1956, p 385-408
29. H.G. Baron and B.S. Bloomfield, Resistance to Thermal Stress Fatigue of Some Steels, Heat-Resisting Alloys, and Cast Irons, *J. Iron Steel Inst.* (London), Vol 197, Part 3, Mar 1961, p 223-232
30. L.-A. Norstrom, B. Johansson, and N. Ohrberg, "Thermal Fatigue and Thermal Shock Behavior of Martensitic Hot Work Steel," Paper No. G-T81-093, 11th International Die Casting Congress and Expo

(Cleveland, OH), Society of Die Casting Engineers, 1981

31. J. Worbye, New Information Points the Way to Longer Die Casting Die Life, *Die Cast. Eng.,* Vol 29 (No. 4), 1985, p 42
32. J.-A. Gaven, L.-A. Norstrom, and J. Worbye, "Performance of Newly Developed Hot Work Tool Steel for Die Casting," Paper No. G-T83-011, 12th International Die Casting Congress and Expo (Minneapolis), Society of Die Casting Engineers, 1983
33. V.A. Kortesoja, Ed., *Properties and Selections of Tool Materials,* American Society for Metals, 1975, p 271
34. T. Arai and T. Iwama, "Carbide Surface Treatment of Die Cast Dies and Components," Paper No. G-T81-092, 11th International Die Casting Congress and Expo (Cleveland, OH), Society of Die Casting Engineers, 1981
35. M.J. Mirtich, C.-Y. Nieh, and J.F. Wallace, "Sputtered Protective Coatings for Die Casting Dies," NASA-TM-81735, National Aeronautics and Space Administration, 1981
36. V.A. Kortesoja, Ed., *Properties and Selection of Tool Materials,* American Society for Metals, 1975, p 131-170
37. D. Schmoeckel and H. Frontzek, Reduction of Wear on Sheet Metal Forming Tools, *Ann. CIRP,* Vol 35 (No. 1), 1986, p 195-198
38. E.V. Finckenstein and H. Brox, Investigation of Coated Thin Sheet Metal During Forming by Bending, *2nd Conference on Advanced Tech. Plast.* (Stuttgart), 1987, p 233-237
39. K. Takaishi, T. Kondo, S. Sugiyama, and M. Mizutame, Statistical Analysis of Blanking Tool Wear by a Group Method of Data Handling, *1st Conference on Advanced Tech. Plast.* (Tokyo), 1984, p 833-838
40. A. Kannappan, Wear in Forging Dies, *Met. Form.,* Vol 36 (No. 12), Dec 1969, p 335; Vol 37 (No. 1), Jan 1970, p 6
41. J.L. Aston and E.A. Barry, A Further Consideration of Factors Affecting the Life of Drop Forging Dies, *J. Iron Steel Inst.,* Vol 210 (No. 7), July 1972, p 520
42. S.M.J. Ali, B.W. Rooks, and S.A. Tobias, The Effect of Dwell Time on Die Wear in High Speed Hot Forging, *Proc. Inst. Mech. Eng.,* Vol 185, 1970-1971, p 1171
43. A. Thomas, Wear of Drop Forging Dies, *Tribology in Iron and Steel Works,* Iron and Steel Institute, London, 1970, p 135
44. A.K. Singh, B.W. Rooks, and S.A. Tobias, Factors Affecting Die Wear, *Wear,* Vol 25, 1973, p 271
45. B.W. Rooks, The Effect of Die Temperature on Metal Flow and Die Wear During High Speed Hot Forging, *Proceedings of the Fifteenth International MTDR Conference* (Birmingham, England), MacMillan, Sept 1974, p 4
46. J.L. Aston, A.D. Hopkins, and K.E. Kirkham, The Wear Testing of Hot Work Die Steels, *Metall. Met. Form.,* Vol 39 (No. 2), 1972, p 46
47. T.M. Silva and T.A. Dean, Wear in Drop Forging Dies, *Proceedings of the Fifteenth International MTDR Conference* (Birmingham, England), MacMillan, Sept 1971, p 22
48. E. Doege, R. Melching, and G. Kowallick, Investigation Into the Behavior of Lubricants and the Wear Resistance of Die Materials in Hot and Warm Forging, *J. Mech. Work. Technol.,* Vol 2, 1978, p 129
49. F.T. Netthöfel, "Contributions to the Knowledge on Wear in Die Materials," Doctoral Dissertation, Technical University of Hannover (in German), 1965
50. H. Hecht and H.M. Hiller, Performance Comparison of Some German and American Die Steels, *Werkstattstechnik und Machinenbau,* Vol 49 (in German), 1959, p 645
51. A. Misra and I. Finnie, A Review of the Abrasive Wear of Metals, *J. Eng. Mat. Technol. (Trans. ASME),* Vol 104, Apr 1982, p 94
52. A. Thomas, Variability of Life in Drop Forging Dies, *Met. Form.,* Vol 38 (No. 2), Feb 1971, p 41
53. C.W. Vigor and J.W. Hornaday, A Thermocouple for Measurement of Temperature Transients in Forging Dies, *Temperature, Its Measurement and Control,* Vol 3, Part 2, Rheinhold, 1961, p 265
54. C. Miland and W. Panasiuk, Increasing the Life of Forging Tools: New Materials, Technologies and

Methods of Investigation, *J. Mech. Work. Technol.,* Vol 6, 1982, p 183

55. F.A. Still and J.K. Dennis, Electrodeposited Wear-Resistant Coatings for Hot Forging Dies, *Metall. Met. Form.,* Vol 44 (No. 1), 1977, p 10
56. F. Neuberger *et al.,* Increasing Die Life by Surface Welding, *Fertigungstechnik und Betrieb,* Vol 12 (in German), 1962, p 822
57. T.M. Wu, Investigation of Surface Welding of Dies for Forging Steel Parts, *Forschungsbericht des Landes Nordrhein-Westfalen,* No. 1349, Westdeutscher Verlag, Koln, (in German), 1964
58. A. Gray, Dies Hardfaced with Alloy Last Longer, *Iron Age,* Vol 167, May 1951, p 68
59. A New Face Toughens Forge Dies, *Weld. Eng.,* Vol 40, Oct 1955, p 60
60. P.H. Thornton, Effect of Spark Hardening on Life of Hot Forging Dies, *Met. Technol.,* Vol 7, 1980, p 26
61. M. Gierzynska-Dolna, Effect of the Surface Layer in Increasing the Life of Tools for Plastic Working, *J. Mech. Work. Technol.,* Vol 6, 1982, p 193
62. T.A. Dean and C.E.N. Sturgess, Warm-Forming Practice, *J. Mech. Work. Technol.,* Vol 2, 1978, p 255
63. G. Sodero and M. Remondino, "Nitriding of Press-Forging Dies Boosts Output," *Seventh International Drop Forging Conference* (Brussels), Sept 1971
64. M.A. Kellow *et al.,* The Measurement of Temperatures in Forging Dies, *Int. J. Mach. Tool Des. and Res.,* Vol 9, 1969, p 239
65. P. Voss, "Die Temperatures and Die Wear During Forging in an Automatic Forging Machine," Doctoral Dissertation, Technical University of Hannover (in German), 1968
66. B.W. Rooks, A.K. Singh, and S.A. Tobias, Temperature Effects in Hot Forging Dies. *Met. Technol.,* Vol 1 (No. 10), Oct 1974, p 449
67. R.E. Okell and F. Wolstencroft, Suggested Mechanism of Hot Forging Die Failure, *Met. Form.,* Vol 35 (No. 2), Feb 1968, p 41
68. A. Thomas, Cracking of Forging Dies, *Proceedings of the Tenth International Drop Forging Conference* (London), Vol 22, June 1980
69. A. Thomas, Cracking and Fracture of Hot-Work Die Steels, *Proceedings of the Fifteenth International Machine Tool Design and Research Conference* (Birmingham, England), Vol 51, MacMillan, Sept 1974
70. H.Y. Oei, Adhesion Strength of Phosphate Coatings in Cold Forming, *Second Conference on Advanced Tech. of Plasticity* (Stuttgart), Vol II, 1987, p 893-899
71. Cold Extruder Uses MoS_2 for Economy, *Molysulfide Newsletter,* Climax Molybdenum Co., Vol XII (No. 3), Apr 1969
72. R.M. Davidson and T.L. Gilbert, Additions of MoS_2 to a Bonded Solid Lubricant for Severe Ironing Application, *Wear,* Vol 31, 1975, p 173-178
73. H.M. Schiefer, G.V. Kubczak, and W. Laepple, Hard Coating for Metal Forming, *Metalworking Lubrication,* S. Kalpakjian and S.C. Jain, Ed., American Society of Mechanical Engineers, 1980
74. N.P. Suh, *Tribophysics,* Prentice-Hall, 1986, p 454-487
75. T.A. Dean and C.E.N. Sturgess, Warm-Forming Practice, *J. Mech. Work. Technol.,* Vol 2, 1987, p 255
76. G. Sodero and M. Remondino, Nitriding of Press-Forging Dies Boosts Output, *Seventh International Forging Conference* (Brussels), Sept 1971
77. A. Vincze, "Surface Hardening of Forging Dies by Boriding," *Bányász. Kohász. Lapok (Kohász.),* Vol 102 (No. 11) (in Hungarian), Nov 1969, p 480
78. V.S. Burgreev and S.A. Dobnar, Electrolytic Boriding of Hammer Forging Dies and Their Heat Treatment, *Met. Sci. Heat Treat.,* Vol 14 (No. 6), June 1972, p 513
79. H.C. Fiedler and R.J. Sieraski, Boriding Steels for Wear Resistance, *Met. Prog.,* Vol 99 (No. 2), Feb 1971, p 101
80. H.C. Child, The Heat Treatment of Tools and Dies--A Review of Present Status and Future Trends, *Tools and Dies for Industry,* The Metals Society, London, 1977
81. J.K. Hirvonen, Surface Modification of Polymers and Ceramics, *Adv. Mater. Proc.,* May 1986

82. N.E.W. Hartley, W.E. Swindlehurst, E. Dearnaley, and J.F. Turner, *J. Mater. Sci.,* Vol 8, 1973, p 900-904
83. J.K. Hirvonen, *Proceedings of the Materials Research Society Symposium on Ion Implantation and Ion Beam Processing of Materials,* Materials Research Society, 1983
84. J.K. Hirvonen and C.R. Clayton, *NATO Conf.,* No. 8, 1983, p 323
85. G. Dearnaley, *J. Met.,* Vol 34, 1982, p 18-27
86. M. Iwaki, *Thin Solid Films,* Vol 101, 1983, p 223-231
87. H. Hearman, *Nucl. Instrum. Methods,* Vol 182/183, 1981, p 887-898
88. D.I. Potter, M. Ahmed, and S. Lamond, *J. Met.,* Vol 35, 1983
89. I.L. Singer, *Proceedings of the Materials Research Society Symposium on Ion Implantation and Ion Beam Processing of Materials,* Materials Research Society, 1983
90. R.E. Hoisington, Extending Tool Life with Ion Implantation, *G.E. Technol.,* Jan 1986, p 9-12
91. R.E. Hoisington, Multiplying Die Life Through Ion Implantation, *Met. Stamp.,* June 1987, p 3-6
92. J. Hohisha, Ion Beams and Metals, *The New York Times,* 23 Oct 1986, p 30
93. G. Dearnaley, "The Treatment of Metal-Working Tools by Ion Implantation," Paper No. C 142/80, Institute of Mechanical Engineers, London, 1980
94. H. Hayashi, K. Kawashima, J. Fujihana, and M. Iwaki, Effects of Ion Implantation on Mechanical Properties of Metal and Its Application to Plastic Working, *Adv. Technol. Plast.,* Vol I, Springer-Verlag, Berlin, 1987, p 335-341
95. C. Weist, G.K. Wolf, and P. Ballhause, Improving Wear Resistance of Metal Forming Tools by Ion Implantation, to be published in *Mater. Sci. Eng.*
96. J.R. Conrad, "Method and Apparatus for Plasma Source Ion Implantation," Application to U.S. Patent and Trademark Office, 20 Jan 1987
97. J.R. Conrad and C. Forest, *IEEE International Conference on Plasma Sciences* (Saskatoon, Canada), Paper 2D7, 19-21 May 1986
98. J.R. Conrad and T. Castagna, *Bull. Am. Phys. Soc.,* Vol 31, 1986, p 1474
99. F.J. Worzala, R.A. Dodd, J.R. Conrad, and R. Radike, *Proceedings of the International Conference on Tool Materials* (St. Charles, IL), 28 Sept-1 Oct 1987, sponsored by Uddeholm Research Foundations, Colorado School of Mines Press
100. T. Arai, Carbide Coating Process by Use of Molten Borax Baths in Japan, *J. Heat Treat.,* Vol 1 (No. 2), 1981, p 15-22
101. "MetalLife: Extended Life and Improved Performance from Dies," Brochure, Badger Metal Tech., Inc.
102. "MetalLife: Extended Life and Improved Performance for Die Casting," Brochure, Badger Metal Tech., Inc.
103. J.V. Skoff, MetalLife for Lubricant Retention; A Key to Better Tool Performance, *Met. Stamp.,* Sept 1986, p 8-12
104. Micro-Precision Peening Process Boosts Die Life, *Precis. Met.,* Dec 1986, p 9
105. E.S. Hamel, "Ceramic Coatings: More Than Just Wear Resistant," *Mech. Eng.,* Aug 1986, p 30-86
106. R. Bonetti, Hard Coatings for Improved Tool Life, *Met. Prog.,* Vol 119 (No. 7), June 1981, p 4
107. M. Podob, Scientific Coatings, Inc., private communication, 1981
108. P.H. Thornton and R.G. Davies, High-Rate Spark Hardening of Hot-Forging Dies, *Met. Technol.,* Vol 6, 1979, p 69
109. P.H. Thornton, Effect of Spark Hardening on Life of Hot-Forging Dies, *Met. Technol.,* Vol 7, 1980, p 26
110. H.G. Sanborn, F. Carago, and J.R. Kreeger, "Refractory Metal Coated Metal-Working Dies," U.S. patent 4,571,983, 25 Feb 1986
111. *Heat Treating, Cleaning and Finishing,* Vol 2, 8th ed., *Metals Handbook,* American Society for Metals, 1964
112. S.L. Scheier and R.E. Christin, Drop Forge Dies--Hard Chromium Plating Cuts Cost of Die Sinking, *Met. Prog.,* Vol 56 (No. 4), Oct 1949, p 492

113. F.A. Still and J.K. Dennis, The Use of Electrodeposited Cobalt Alloy Coatings to Improve the Life of Hot Forging Dies, *Electroplat. Met. Finish.,* Vol 27 (No. 9), Sept 1974, p 9

114. K.J. Lodge *et al.,* The Application of Brush-Plated Cobalt Alloy Coatings to Hot and Cold-Work Dies, *J. Mech. Work. Technol.,* Vol 3, 1979, p 63

115. F.A. Still and J.K. Dennis, Electrodeposited Wear-Resistant Coatings for Hot-Forging Dies, *Metall. Met. Form.,* Vol 44 (No. 1), 1977, p 10

116. J.K. Dennis and F.A. Still, The Use of Electrodeposited Cobalt Alloy Coatings to Enhance the Wear Resistance of Hot Forging Dies, *Cobalt,* Vol 1, 1975, p 17

117. J.K. Dennis and D. Jones, Brush Plated Cobalt Molybdenum and Cobalt-Tungsten Alloys for Wear Resistant Applications, *Tribol. Int.,* Vol 14, 1981, p 17

118. H. Moestue, "Hardfacing and Reclamation of Hot Work Dies," unpublished manuscript, Sveiseindustri, Oslo, 1976

119. O. Knotek, Wear Prevention, *Fundamentals of Tribology,* N.P. Suh and N. Saka, Ed., MIT Press, 1978, p 927-941

120. O. Knotek, E. Lugscheider, and H. Reiman, Wear Resistance Properties of Sprayed and Welded Layers, *Metallurgical Aspects of Wear,* Deutsche Gessellshaft für Metallkunde, E.V. (DGM), 1979, p 307-318

121. J. Gunser, Wall Colmonoy Corp., private communication, 1981

122. G.L. Sheldon and R.N. Johnson, Electrospark Deposition--A Technique for Producing Wear Resistant Coatings, *Wear of Materials 1985,* K.C. Ludema, Ed., American Society of Mechanical Engineers, 1986, p 338-396

123. C. Weist and C.W. Westheide, Application of Chemical and Physical Methods for the Reduction of Tool Wear in Bulk Metal Forming Processes, *Ann. CIRP,* Vol 35 (No. 1), 1986, p 199-204

124. M. Gierzynska-Dolna, Effect of the Surface Layer in Increasing the Life of Tools for Plastic Working, *J. Mech. Work. Technol.,* No. 6, 1982, p 193-204

Friction and Wear in the Mining and Mineral Industries

D.L. Olson and C.E. Cross, Center for Welding and Joining Research, Colorado School of Mines

Introduction

THE PRACTICE of mining and mineral processing, by its nature, involves severe mechanical interactions between metals, as well as between abrasive nonmetallic and metallic materials. The abrasive nature of most ores can cause significant wear to both handling and processing equipment.

Various wear processes are involved, depending on the nature of the abrasive material, type of loading, and the environment. The direct impact of metallic components with the earth requires alloys that have excellent ductility and work-hardening properties. The transport of ground material across a surface, such as chutes, requires high surface hardness. The crushing or fragmentation of ore requires even higher hardness values. Performing mechanical operations in corrosive environments, such as mineral slurry pumps, requires that the material exhibit both corrosion and wear resistance.

To properly select wear-resistant materials, careful analysis of the material application is necessary. This selection should consider a very broad range of alloys, including high-strength steels, white irons, austenitic alloys, and high-chromium/high-carbon stainless steels.

Types of Damage

Abrasive wear, the most common form of damage in the mineral industry, can be classified as being due to low-stress, high-stress, or impact abrasion (Ref 1, 2). The material-ore interaction is considered a low-stress abrasion application when the wear process does not involve a fracture of the abrasive material.

An example is the movement of fragments of ore traveling across a high-strength steel surface (Fig. 1a). Low-stress abrasion is not accompanied by significant impact. The damage is the result of either scratching or a micrometal removal process. Sharp, angular abrasives result in the highest wear rate. For this case, materials with high hardness are used to minimize the metal removal rate. Abrasion-resistant (AR) steels, alloy cast irons, and ceramic tiles are often used for these applications.

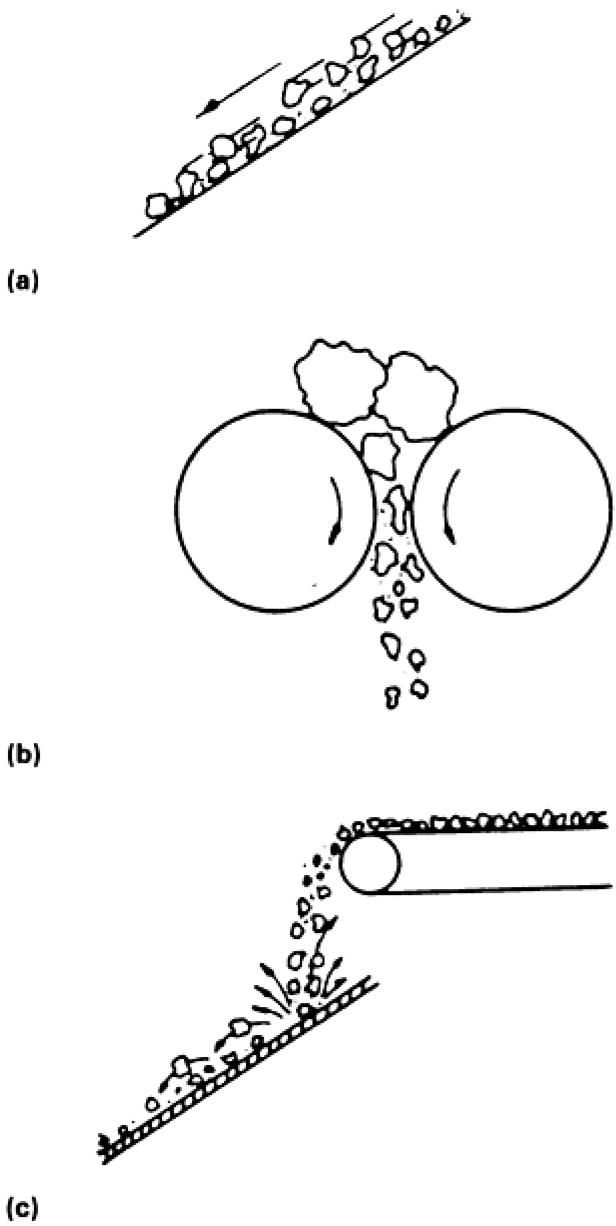


Fig. 1 Three types of abrasive wear, (a) Low stress. (b) High stress. (c) Impact. Source: Ref 1

High-stress abrasion results when abrasive materials are caught between two normally loaded surfaces. The loads are commonly sufficient not only to cause particle penetration of the loading surfaces, but can also fragment the abrasive material. This damage is common to ball mills, rock drills, and rock crushers (Fig. 1b). Surface damage is due to a combination of scratching and plastic deformation, commonly resulting from fatigue. The damage can be minimized by the selection of materials with both high yield strength and hardness.

Some sites on mining equipment will be subject to direct impact, which requires those sites to have both hardness and toughness. The high localized pressures on impact cause the abrasive materials to cut into the wearing surface, resulting in very apparent large gouges and scratches. This type of damage (Fig. 1c) is common to the impact areas of chutes, pulverizing mills, and shovels (Ref 1). The material of choice is often an austenitic alloy that has high toughness and progressively increase in hardness with use through work hardening (Ref 3, 4). High-manganese steels with greater than 14 wt% Mn (known as Hadfield alloys) are examples of traditional impact-resistant austenitic alloys.

Corrosive Wear. Wear is primarily the removal of surface metal by mechanical action. Often, the mining and mineral processing environment is wet, which cause a chemical action that affects the mechanical wear (Ref 5 , 6, 7, 8, 9). This corrosion effect either increases or decreases the metal removal rate, depending on the natures of both the mechanical

interaction and the corrosion product. The synergistic effect of mechanical and chemical interaction causes considerable variation in reported wear rates on mining and mineral processing equipment.

Adhesive wear between one metallic part and another is also a common form of damage in the mineral industry. It results from the surface tearing of materials that became bonded because of rubbing (Ref 1). The damage involves scoring, galling, or seizure. The adhesive wear rate is usually larger when the interacting materials have similar compositions and crystal structures, and it usually occurs under load with little or no lubrication. A common location for adhesive wear is at a surface where wire rope rubs against spools and pulleys, for example. Figure 2 illustrates adhesive wear on a steel spool that has been interacting with wire rope.

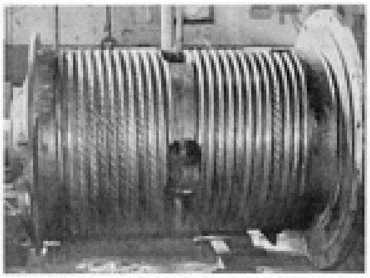


Fig. 2 Cable wear on large spool, a form of adhesive wear. Source: Ref 10

Testing for the types of wear damage described above is often performed according to ASTM test procedures (Ref 11, 12, 13), as well as other special tests designed by specific corporations to evaluate special conditions (Ref 14, 15, 16, 17).

Methods to Improve Wear Resistance

Wear resistance can be increased on mining and mineral processing equipment through proper material selection and equipment design, and through the use of wear plates, as well as hardfacing deposition. The method selected will depend on the type and use of the part and the economics associated with operation of the equipment.

Material Selection. A variety of materials that provide wear resistance are commercially available (Ref 18, 19). Some material types are shown in Fig. 3 relative to service and environmental needs.

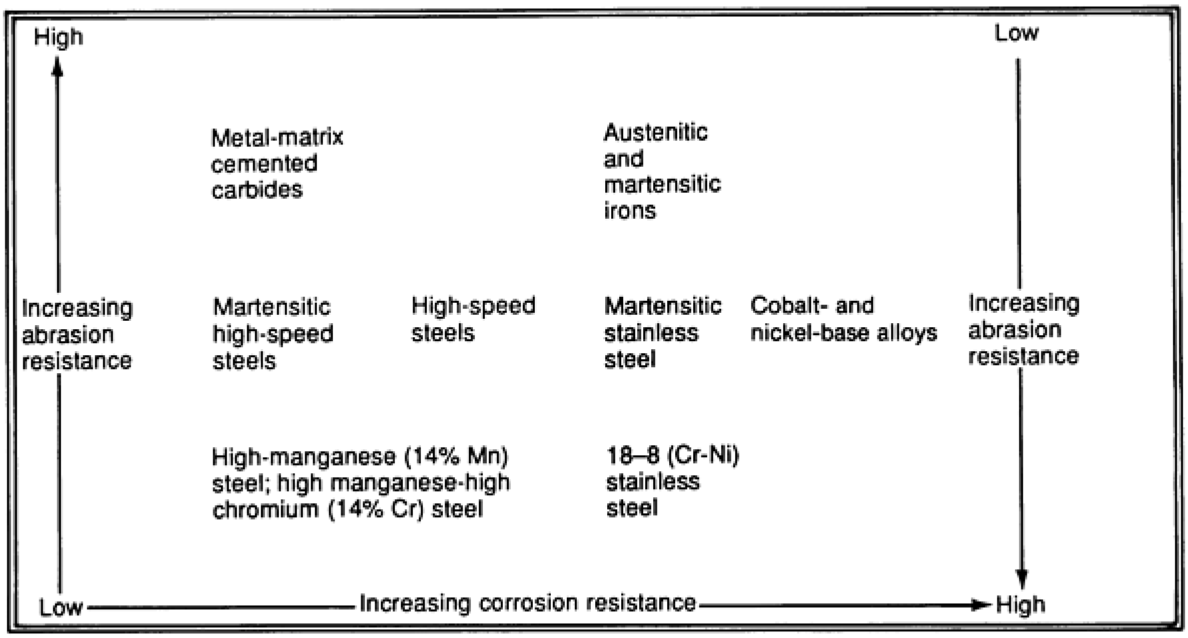


Fig. 3 Relative placement of various wear-resistant alloy systems when considering abrasion, impact, and corrosion resistance. Source: Ref 20

High impact resistance with low abrasion and corrosion resistance can be achieved with high manganese ferrous alloys. A Hadfield alloy (14Mn-1C) is a traditional material for impact resistance. In both the as-cast and as-welded conditions, the Hadfield alloy shows good ductility. Additionally, the alloy work hardens with impact and deformation, resulting in increased strength and hardness levels (Ref 21). For more corrosive environments, impact resistance can be achieved with the 18-8 (18%Cr + 8%Ni) stainless steels. Austenitic stainless steels, such as the 18-8's, also work harden with service, but have the disadvantage of high material cost.

Abrasion-resistant plate steels, similar to armor-grade steel, have a unique combination of desirable properties: high hardness, good toughness, and weldability (Ref 22). Abrasion resistance is achieved using a water quenched and tempered martensite, which is often low enough in alloy content (that is, has a low carbon equivalent) to avoid the concerns of weld underbead cracking with proper welding practice. Vacuum degassing and desulfurization during steelmaking is required to minimize inclusions and thus improve toughness in the short transverse direction.

As the abrasion resistance increases from a low-stress to high-stress application, the selected material must increase in hardness from martensitic alloy steel to high-carbide materials, including alloyed white irons. Applications for martensitic steels of various hardness levels are given in Table 1.

Table 1 Materials for surfacing, build-up, and hardfacing

Types of Materials	Advantages	Applications
Low hardness		
Pearlitic steels	Low cost, crack resistant	A build-up material to restore dimension or as a low-cost base for surfacing applications
Mild steels	An excellent base for hardfacing	
Low-alloy steels	Some toughness for build-up of worn areas	
Austenitic steels	Excellent for heavy impact, tough	For metal-to-metal wear under heavy impact conditions
Low-carbon Ni-Cr stainless	Good corrosion resistance	Corrosion-resistant surfacing of large tanks, nuclear vessels
High-carbon Ni-Cr stainless	Oxidation and hot wear resistant	Furnace parts, red heat frictional wear
14% Mn, Cr-Ni	High yield strength austenitic	Build-up, crack repair, joining manganese alloys or manganese to mild steel
14% Mn, 1% Mo	Fair abrasion and corrosion resistance, work	Railways track build-up

	hardens	
14% Mn, 3% Ni	Work hardens	Build-up and joining of manganese parts
Semiaustenitic steels	Low-cost crack-resistant materials	Hardfacing applications
Copper alloys	Resistance to seizing under frictional wear conditions	For build-up of bearing surfaces
Medium hardness		
Nickel-base alloys	Good corrosion resistance and excellent hot hardness	For applications where hot hardness is important
Nickel-chromium	Resistance to oxidation	For corrosive wear conditions
Nickel-chromium-molybdenum	Resistance to exhaust gas corrosion	For hardfacing truck, bus, and aircraft engine valves
Nickel-chromium-molybdenum tungsten	Corrosion-resistant properties	For surfacing applications
Nickel-chromium-boron	Corrosion-resistant as well as abrasive-resistant properties	Pumping service applications, especially in the oil fields
Martensitic alloy irons	Excellent abrasion resistance	For overcoming abrasive wear where light or no impact is involved
Chromium-molybdenum	Abrasion resistance while withstanding light impact	For the build-up of machine parts subject to repetitive metal-to-metal wear with or without light impact
Chromium-tungsten	High compressive strengths	
Austenitic alloy irons	Better crack-resistant properties than the martensitic alloy irons	Where erosive wear with or without light impact is present
Chromium-molybdenum	Good crack-resistant qualities	Pump and turbine build-up
Nickel-chromium	Crack resistant, light impact	For erosive wear with light impact
Martensitic steels	Abrasion resistance with medium impact	For a variety of abrasive wear conditions where there may also be involvement with medium impact
Low carbon (up to 0.30% C)	Least expensive, tough	
Medium carbon (0.30-0.65% C)	Good resistance to medium impact	
High carbon (0.65-1.7% C)	Fair abrasion resistance	
High hardness		
High-chromium irons	Excellent erosion-resistant properties	For general use where hot gases or materials are involved
Austenitic	Good abrasion-resistant qualities	For farm and earth-moving equipment working in sandy soils
Martensitic	Ability to be rehardened after annealing	As a hardfacing in steel mills and refineries where hot erosion (595 °C, or 1100 °F) is a wear problem
Tungsten-molybdenum alloys	High red hardness	For hardfacing of coke oven parts as well as other hot (425-650 °C, or 800-1200 °F) steel mill applications
Chromium-tungsten cobalt alloys	High red hardness, high hot strength and creep-resistant qualities	For most hardfacing applications, but best suited to those applications where hot wear and abrasion (above 650 °C, or 1200 °F) are involved (jet engine turbine blades, gas engine exhaust valve)
Low carbon (up to 1.0% C)	Good oxidation resistance, tough	
Medium carbon (1.4% C)	Resistant to abrasion and oxidation	
High carbon (2.5% C)	Good abrasion resistance, brittle	
Tungsten carbides	The ultimate in abrasion-resistant qualities	A variety of materials used for hardfacing to meet an extremely wide range of severe abrasive conditions, especially oil well drill bits, tool joints, rock drill bits
Fine-grain tubular rods	A relatively smooth, extremely wear-resistant surface, gas weld	
Coarse granular tube rod	Develops a rough cutting surface as the deposited binder wears away	
Granules or inserts	Develops an extremely rough abrasive cutting surface as granules or inserts become more exposed. Highly abrasive, but brittle under impact conditions	

Source: Ref 23

The role of carbon content in steel and iron to promote abrasion resistance is illustrated in Fig. 4, which indicates the position of various alloy types relative to their gouging wear rates. Figure 5 correlates wear rate to hardness for the various alloys experiencing abrasion wear.

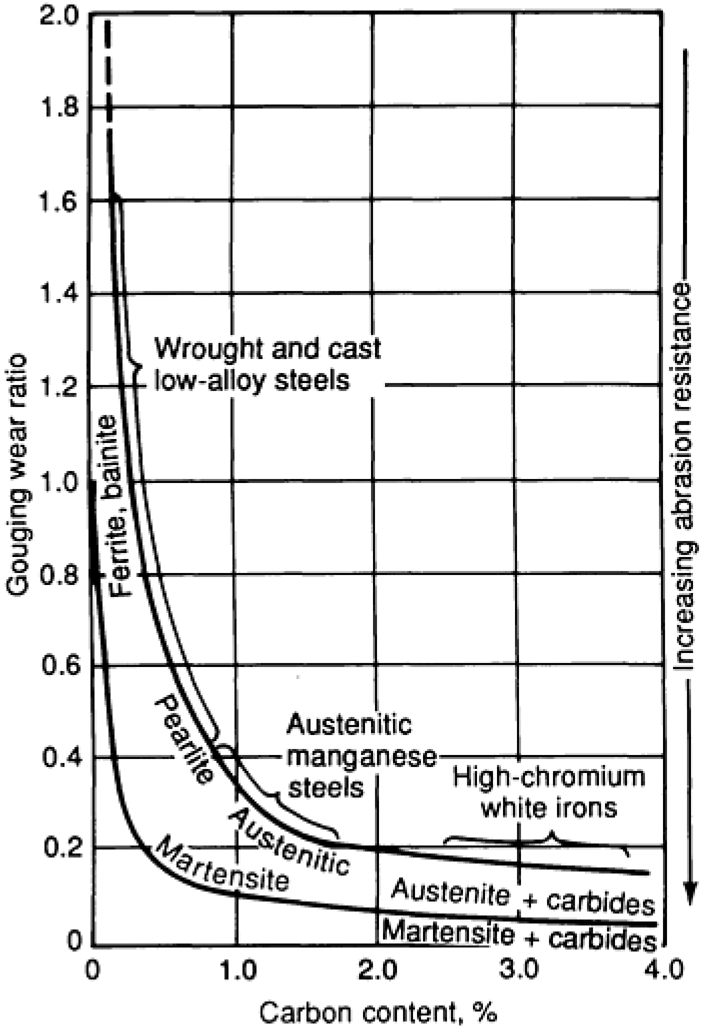


Fig. 4 Relative placement of various alloys as function of carbon content and abrasion resistance, as measured by gouging wear ratio. Source: Ref 24

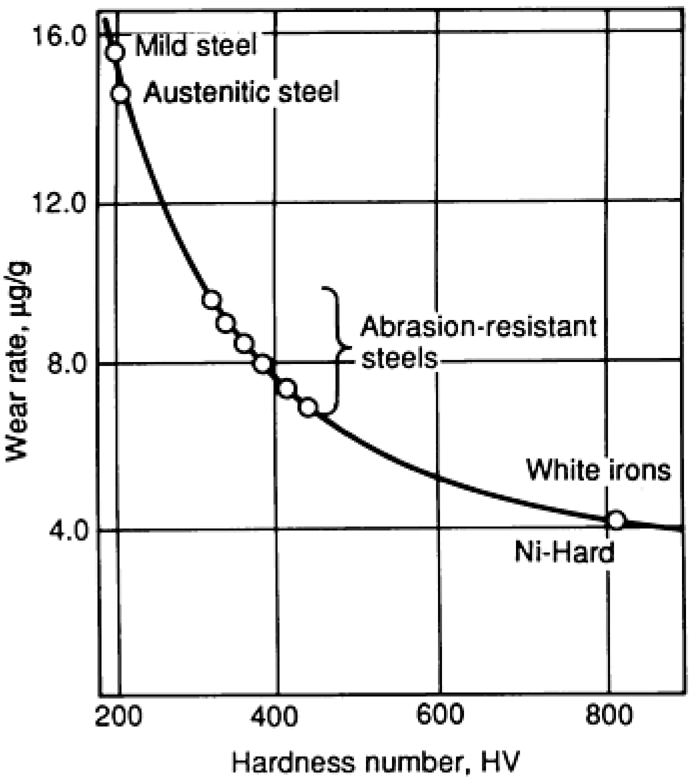


Fig. 5 Abrasion resistance of various alloy systems, comparing wear rate to hardness. Source: Ref 25

Use of Wear Plates. A part that experiences major wear should be designed for easy replacement to prevent excessive processing downtime. Figure 6 shows high-manganese steel wear plate castings that are welded into the bucket of a large shovel. These castings were made with properly spaced holes, which alloys easy shielded metal are welding of the cast plates to the shovel. Similar wear plates are used as working surface in crushers (Ref 26, 27) and pulverizers.

Fig. 6 High-manganese steel-casted wear plates plug welded to inside of large shovel. Source: Ref 10

Figure 7 illustrates the use of wear plates in medium-sized buckets. Often, an easily weldable steel plate or strip that is considered expendable is used. A typical practice is to incur significant wear in a new bucket to reduce its weight prior to welding these wear plates to it. Bucket weight is an important consideration in the efficiency and economics of earth-moving processes.

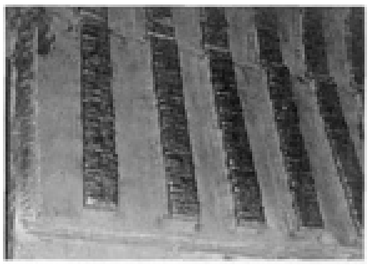


Fig. 7 Steel strips welded onto surface of shovel and hardfacing deposition placed on top of steel strips to give abrasion and impact resistance for a rack consisting of large racks. Source: Ref 10

Table 2 lists the various ferrous alloys used for mill liners. They include a broad range of high-carbon steel and alloyed white irons.

Table 2 Ferrous materials for grinding mill liners

Item	Material	Composition range, %(a)							Hardness range, HB(b)	Relative wear rate(c)
		C	**Mn**	**Si**	**Cr**	**Mo**	**Ni**	**Cu**		
1	Martensitic Cr-Mo white iron	2.4-3.2	0.5-1.0	0.5-1.0	14.0-23.0	1.0-3.0	0-1.5	0-1.2	620-740	88-90
2	Martensitic high-carbon Cr-Mo steel	0.7-1.2	0.3-1.0	0.4-0.9	1.3-7.0	0.4-1.2	0-1.5	. . .	500-630	100-111(d)
3	Martensitic high-Cr white iron	2.3-2.8	0.5-1.5	0.8-1.2	23.0-28.0	0-0.6	0-1.2	. . .	550-650	98-100
4	Martensitic Ni-Cr white iron	2.5-3.6	0.3-0.8	0.3-0.8	1.4-2.5	0-1.0	3.0-5.0	. . .	520-650	105-109
5	Martensitic medium-carbon Cr-Mo steel	0.4-0.7	0.6-1.5	0.6-1.5	0.9-2.2	0.2-0.7	0-1.5	. . .	500-620	110-120
6	Austenitic 6Mn-1Mo steel	1.1-1.3	5.5-6.7	0.4-0.7	0.5 max	0.9-1.1	. . .	. . .	190-230	114-120
7	Pearlitic high-carbon Cr-Mo steel	0.5-1.0	0.6-0.9	0.3-0.8	1.5-2.5	0.3-0.5	0-1.0	. . .	250-420	126-130
8	Austenitic 12Mn steel	1.1-1.4	11.0-14.0	0.4-1.0	0-2.0	0-1.0	. . .	. . .	180-220	136-142
9	Pearlitic high-carbon steel	0.6-1.0	0.3-1.0	0.2-0.4	. . .	. . .	. . .	. . .	240-300	145-160
10	Pearlitic white iron	2.8-3.5	0.3-1.0	0.3-0.8	0-3.0	. . .	. . .	. . .	370-530	ND(e)

Source: Ref 26

(a) The composition range from an individual supplier will normally be narrower than the ranges given in the table. Production of certain compositions falling within the above ranges may involve proprietary rights covered by patents or trademarks.

(b) Hardnesses listed are on the unworn surface of the liners. Austenitic and martensitic alloys tend to work harden on wearing surfaces.

(c) Relative wear rate as determined when wet grinding minus 10mm ($\frac{3}{8}$ in.) feed to minus 48 mesh on ore containing about 65% quartz, 25-30% feldspar, and 3% pyrite as the principal abrasives, in primary ball mills at Climax, Colorado.

(d) A composition containing about 1.0% C, 5-6% Cr, and 1.0% Mo was used as a comparative standard in all tests and assigned a relative wear rate of 100.

(e) Wear rates in the Climax mills could not be determined on this material, because of spalling and breakage.

Use of Hardfacing Deposition. Three primary methods are used to deposit hardfacing materials in the mineral industry (Ref 28, 29, 30, 31, 32, 33, 34, 35): arc welding, oxy-fuel gas welding, and thermal spraying. Filler metals are available in rod, wire, and powder form.

The most traditional process is shielded metal arc (SMA) weld deposition, which involves low capital investment, can be used in all positions, and is an excellent field welding process. It is considered a low deposition rate process. A large selection of commercial electrode compositions exist.

The gas-metal arc welding process is used to make wear-resistant depositions. The gas-metal arc wire electrode process increases deposition rate and productivity, but usually requires more equipment setup. This gas-metal arc weld deposition process is more flexible than the submerged arc welding process, but it is more difficult than shielded metal arc welding in terms of moving the welding equipment, including the necessary gas supply for protective cover.

The submerged arc welding process deposits in either the flat position or on top of a rotating round part and is capable of high deposition rates. It is a fully automatic process that needs to be performed in a stop. A wide range of alloy consumables are available, including strip electrodes.

Flux-cored electrodes that introduce flux and carbide-forming ingredients to the weld deposit, by additions to the core of the welding wire, are available. These flux core additions can also contain generators of plasma and shielding gas which can produce a wire welding process that is self shielding and therefore does not require additional supplied shielding gas.

Hardness is commonly used as a measure in the selection of hardfacing materials. The abrasion resistance of martensitic alloys is expected to increase with increasing hardness, whereas impact resistance would be obtained from materials of low hardness. An extensive list of hardfacing and surfacing materials is given in Table 1, along with their specific advantages. The alloy compositions and AWS designations for these various hardfacing materials are given in Table 3.

Table 3 Classification and Composition of typical hardfacing alloys

Group and description	Typical composition, %										Hardness, HV	AWS specification A5.13-70
	Fe	C	Cr	Mn	Mo	V	W	Co	Ni	B		
1. Iron-base containing less than 20% alloy additions												
Carbon steel	bal	0.5	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	250	. . .
Martensitic alloy steel	bal	0.1	1	0.7	. . .	. . .	. . .	. . .	. . .	. . .	250	. . .
Martensitic alloy steel	bal	0.1	3.5	1	. . .	. . .	. . .	. . .	. . .	. . .	350	. . .
Martensitic alloy steel	bal	0.25	1	1	0.3	. . .	. . .	. . .	. . .	. . .	350	. . .
Martensitic alloy steel	bal	0.35	3	1	. . .	. . .	. . .	. . .	. . .	. . .	450	. . .
Martensitic alloy steel	bal	0.45	5	1	5	0.75	. . .	. . .	. . .	. . .	650	. . .
Martensitic stainless steel	bal	0.1	12	. . .	. . .	. . .	. . .	. . .	. . .	. . .	400	. . .
Martensitic stainless steel	bal	0.25	13	. . .	. . .	. . .	. . .	. . .	. . .	. . .	450	. . .
High-speed steel	bal	0.8	4	0.5	5	2	6	. . .	. . .	. . .	650	Fe 5-A[(a)]
High-speed steel	bal	0.7	4	0.5	7	1	1.5	. . .	. . .	. . .	650	Fe 5-B[(a)]
High-speed steel	bal	0.4	4	0.5	7	1	1.5	. . .	. . .	. . .	600	Fe 5-C[(a)]
Austenitic Mn steel	bal	0.7	0.5	14	. . .	. . .	. . .	. . .	4	. . .	600 (max)	Fe Mn-A[(a)]
Austenitic Mn steel	bal	0.7	0.5	14	1	. . .	. . .	. . .	. . .	. . .	600 (max)	Fe Mn-B[(a)]
2. Iron-base containing more than 20% alloy additions												
Austenitic Cr-Mn steel	bal	0.35	14	14	1	0.4	. . .	. . .	. . .	. . .	600 (max)	. . .
High-speed steel	bal	. . .	10	. . .	. . .	. . .	15	. . .	2.5	. . .	750	. . .
Austenitic steel	bal	0.1	18	. . .	3	. . .	. . .	. . .	8	. . .	500 (max)	. . .
Austenitic iron	bal	4	30	6	. . .	. . .	. . .	. . .	. . .	. . .	700	Fe Cr-Al[(a)]
Martensitic iron	bal	2.5	28	1	. . .	. . .	. . .	. . .	. . .	. . .	600	. . .
3. Nonferrous												
Co-Cr-W	3	1	26	1	. . .	. . .	5	bal	3	. . .	400	Co Cr-A
Co-Cr-W	3	1.4	31	1	. . .	. . .	8	bal	3	. . .	500	Co Cr-B
Co-Cr-W	3	2.5	32	1	. . .	. . .	12	bal	3	. . .	630	Co Cr-C
Ni-Cr-B	3	0.4	11	. . .	. . .	. . .	. . .	1	bal	2.5	400	Ni Cr-A
Ni-Cr-B	4	0.6	13	. . .	. . .	. . .	. . .	1	bal	3	530	Ni Cr-B
Ni-Cr-B	4	0.8	15	. . .	. . .	. . .	. . .	1	bal	4	720	Ni Cr-C
Ni-Cr-Mo-W	5	0.06	15	. . .	16	. . .	4	. . .	bal	. . .	300	AWS A5.11-69

			E Ni Mo Cr-1
4. Carbides			
Tubular rods	50-60% tungsten carbide granules 40-50% Fe		AWS A5.21-70
		Carbides	WC 20/40
		>1800 HV	WC-40/120 and so forth
		Arc deposit	WC-40/120 and so forth
		>1000 HV	. . .
Sintered rods	50-80% tungsten carbide, up to 10% Cr or Ni, 10-50% Fe	>900 HV	. . .

Source: Ref 18

(a) Also covered by AWS A5.21-70, "Composite Surfacing Welding Rods and Electrodes."

Use of Design. The geometric pattern of hardfacing deposition also has a role minimizing the wear rate. Figures 8, 9, 10, 11, 12, and 13 illustrate various patterns in which hardfacing deposits can be laid. Figure 9 suggests that the flow of fine material be perpendicular to stringer bead deposits, whereas with coarse ore, it is common to align the flow of ore with the stringer bead deposits. The larger rocks tend to knock off the perpendicular beads. Ores with mixed sizes require the spaced weave pattern. The general concept is for the interspaces to be filled with fines, which allow the ore to grind upon ore, thereby reducing metal loss. The spacing between deposits is directly related to the size distribution of the ore.

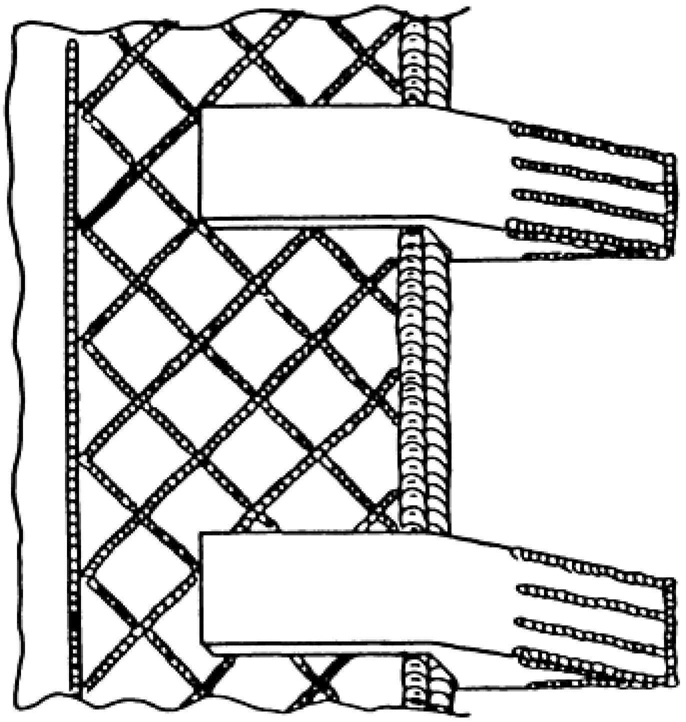


Fig. 8 Typical pattern of hardfacing deposit on lip and teeth of excavator bucket. Complete coverage or hardfacing deposit usually necessary only at edges of lip and teeth. Source: Ref 35

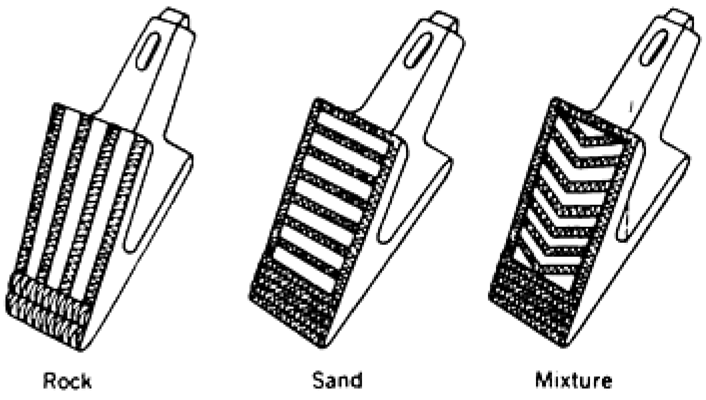


Fig. 9 Different patterns for hardfacing deposits, depending on size and nature of ore material. Source: Ref 32

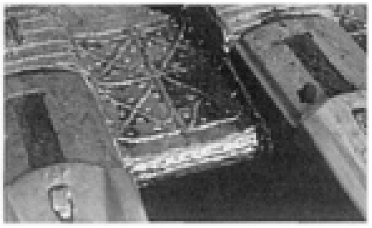


Fig. 10 Hardfacing deposition on lip of shovel that must resist both abrasion and impact. Source: Ref 10

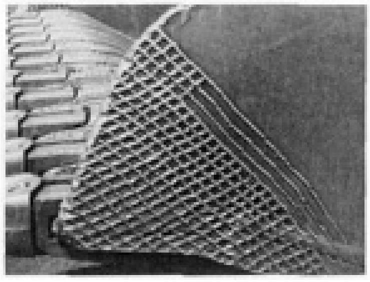


Fig. 11 Typical pattern of hardfacing deposits onto sides of shovels. Source: Ref 10

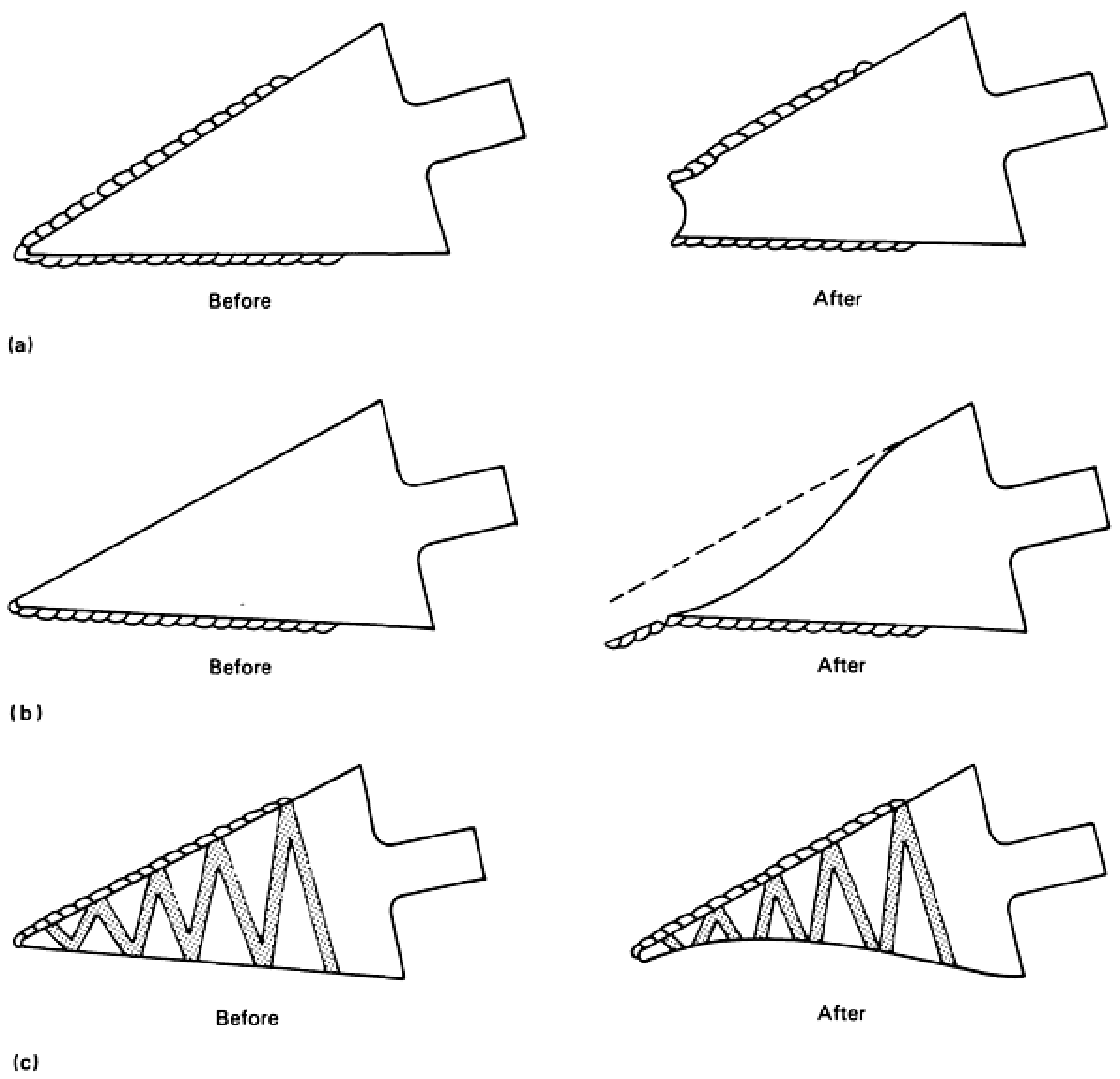


Fig. 12 Effects of various hardfacing coverages on shovel teeth. (a) Hardfacing of both top and bottom teeth (unacceptable). (b) Hardfacing just bottom of tooth (unacceptable). (c) Hardfacing just top of tooth, which is recommended practice that results in self-sharpening of tooth use. Source: Ref 32

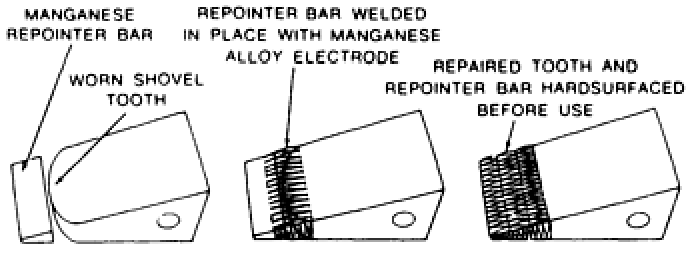


Fig. 13 Worn shovel tooth often rebuilt by welding a new repointer bar, usually made of high-manganese steel, onto end of tooth prior to depositing hardfacing materials. Source: Ref 32

For locations that experience impact loading, it is common to use complete coverage of the hardfacing deposits. Scoop buckets are commonly hardfaced on the digging edge using overlapping stringer beads, with the side of the bucket having a diamond bead pattern, as shown in Fig. 10 and 11.

Special rules are used for hardfacing deposits on scraper teeth. The teeth become blunt, either without hardfacing or with hardfacing on both the top and bottom of the teeth (Fig. 12). Shovel teeth should be hardfaced when they are new. Hardfacing should be deposited with close spacing, if not overlapped, from the point of the tooth back to 50 mm (2 in.) from the point (Fig. 9). The proper method is to deposit hardfacing material only on the top of the tooth. This practice will allow the tooth to preferentially wear, resulting in a self-sharpening effect (Fig. 12). If the point breaks off, then it is often built up with either steel or 14 wt% manganese steel (Fig. 13).

Hardfacing materials commonly have thermal expansion coefficients that are very different from the substrate material on which they are deposited. This thermal expansion mismatch causes residual stresses. The hardfacing materials commonly have thermal expansion values that are larger than the base structural material, which promotes longitudinal tensile stresses in the deposit.

In many hardfacing materials, it is common to have cross-checking, which is a form of transverse cracking that is desirable because it relieves the stresses produced by thermal expansion mismatch between weld deposit and base metal (Fig. 14a). If cross-checking does not occur during hardfacing, then the combined residual and externally applied stresses can cause hairline cracks under the weld deposit in the heat-affected zone of the base material (Fig. 14b), which can result in spalling during service.

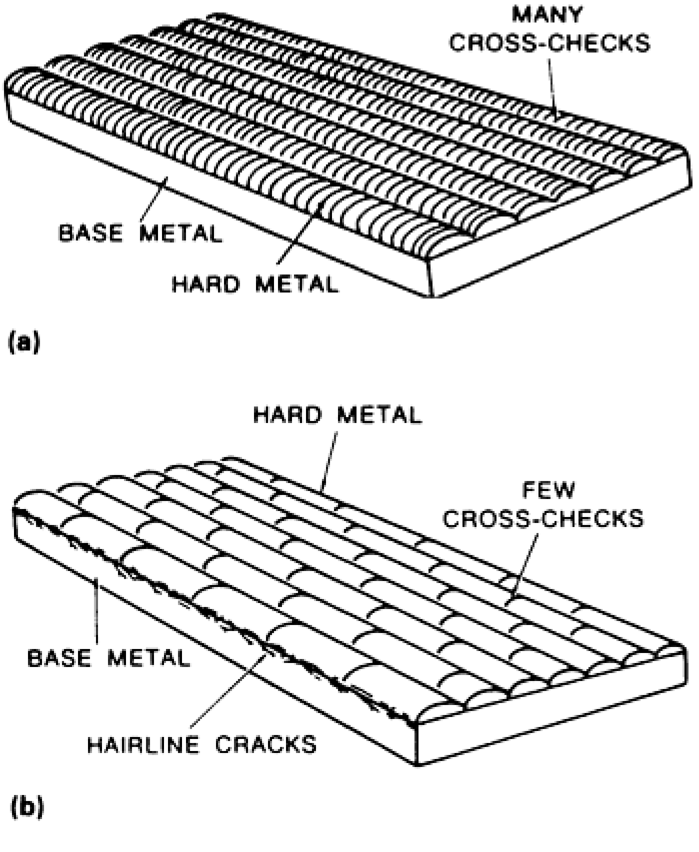


Fig. 14 (a) Sufficient cross-checks. (b) Insufficient cross-checks. Source: Ref 32

In situations where crack-free hard surface deposits are necessary, welding practice that involves a carefully selected welding electrode with preheat and post-heat treatments is necessary (Ref 32). Recommended practices are available from many of the harfacing electrode manufacturers.

Ball and Rod Grinding Media

Grinding media, in the form of balls and rods, are made of various alloys (Ref 3, 24, 36, 37, 38, 39, 40, 41). The mining industry consumes cast steel, Ni-hard, and high-chromium white iron grinding balls. Grinding rods are commonly hot-

rolled high-carbon steel (AISI 1080 to 1095), as well as AISI 52100 steel. The grinding media need to be harder than the ore to be ground. The composition and hardness of typical steel and white iron grinding media are compared in Table 4.

Table 4 Composition and hardness of typical steel and white iron grinding media

Type of ball	Hardness, HB	Composition, %					
		C	**Si**	**Mn**	**Cr**	**Ni**	**Mo**
Forged steel	650	0.80	0.26	0.60	. . .	. . .	0.29
Pearlitic white iron	460	2.75	0.30	0.75	. . .	. . .	. . .
Martensitic white iron	600	3.23	0.62	0.64	1.89	4.26	. . .

Cast iron grinding balls are iron-carbon alloys that contain free carbides. In unalloyed cast iron balls, iron carbide is the phase that influences hardness. With up to 10 wt% chromium, these cast iron balls have mixed carbides, including $(Fe,Cr)_3C$, and hardness values from 800 to 1200 HV. At levels above a 10 wt% chromium addition, $(Fe,Cr)_7C_3$ carbide is also present, resulting in hardnesses that ranges from 1300 to 1800 HV. The nature of these carbides is significant to the wear rate of the grinding media (Ref 38).

Low-alloy pearlitic white iron grinding balls offer a lower-cost alternative to alloyed irons and can be used with many ore types. Pearlitic white irons can have up to 2 wt% chromium content and their hardness values are not extremely high, usually ranging from 450 to 600 HV. They are available in either as-cast or stress-relieved conditions (Ref 38).

Heat-treated martensitic white iron balls are also available. These alloys are usually alloyed with nickel and chromium to allow the formation of a martensitic matrix during the normal cooling process. An example of this type of white iron is Ni-hard cast iron, which has from 2.5 to 4.5 wt% nickel and from 1.5 to 2.5% wt% chromium. Ni-hard cast iron balls have hardness values ranging from 600 to 750 HV, depending on ball size and carbon content (Ref 38).

Martensitic ductile iron balls with hardness values between 600 and 700 HV offer high hardness with higher toughness than is commonly found with pearlitic and martensitic white iron balls. The microstructure consists of a discontinuous carbide network and small graphite nodules in a martensitic matrix. The graphite nodules assist in controlling the amount of free carbides and in promoting the discontinuous carbide network. Martensitic ductile irons are cast to yield carbides and then water quenched to produce a martensitic structure. The toughness can be improved by tempering (Ref 38).

References

1. V.H. Davies and L.A. Bolton, The Mechanism of Wear, *Weld Surfacing and Hardfacing,* The Welding Institute, Cambridge, UK, 1980, p 4-10
2. D.J. Dunn, Metal Removal Mechanisms Comprising Wear in Mineral Processing, *Proceedings of International Conference, Wear of Materials 1985,* American Society of Mechanical Engineers, 1985, p 501-508
3. H.A. Fabert, Jr., Manganese Steels in Cracking and Grinding Service, *Proceedings of Materials for the Industry,* Climax Molybdenum Co., AMAX, 1974, p 163-168
4. J. Tasker, Austenitic Manganese Steel--Fact and Fallacy, *Proceedings of Intermountain Minerals Symposium,* Climax Molybdenum Co., AMAX, 1982 p 3-20
5. W. Day, Corrosion and Erosion Effects on Materials for Slurry Pumps, *Proceedings of Intermountain Minerals Symposium,* Climax Molybdenum Co., AMAX, 1982, p 79-82
6. I. Iwasaki, K.A. Natarajan, S.C. Riemer, and J.N. Orlich, Corrosive and Abrasive Wear in Ore Grinding, *Proceedings of International Conference, Wear of Materials 1985,* American Society of Mechanical Engineers, 1985, p 509-518
7. J.M. Karhnak, "Corrosion and Wear Problems Associated with the Mining and Mineral Processing Industry," Paper 230, Corrosion 84, National Association of Corrosion Engineers, 1984
8. A.I. Asphahani and P. Crook, "Corrosion and Wear of High Performance Alloys in the Mining Industry," Paper 228, Corrosion 84, National Association of Corrosion Engineers, 1984

9. D.J. Singleton and R. Blickenscherfer, "Wear and Corrosion of 12 Alloys During Laboratory Milling of Phosphate Rock in Phosphoric Acid Waste Water," Report RI 9819, U.S. Bureau of Mines, 1984
10. D.L. Olson and W.M. Mueller, "Comprehensive Survey of Material Problems Associated with Welding in the Mining Industry," Final Report G0166160, Colorado School of Mines, 1977, p 1-67
11. Continuous Dry Sand/Rubber Wheel Abrasion Tests, G 65-85, *Annual Book of Standards,* Vol 03.02, ASTM, 1990, p 235-247
12. Standard Test Method for Wet Sand/Rubber Wheel Abrasion Tests, G 105-89, *Annual Book of Standards,* Vol 03.02, ASTM, 1990, p 423-431
13. Standard Practice for Jaw Crusher Gouging Abrasion Test, G 81-83, *Annual Book of Standards,* Vol 03.02, ASTM, p 332-337
14. T.E. Norman and E.R. Hall, Abrasive Wear of Ferrous Materials in Climax Operations, *Evaluation of Wear Testing,* STP 446, ASTM, 1969, p 91-113
15. F. Borik and D.L. Sponseller, Gouging Abrasion Test for Materials Used in Ore and Rock Crushing: Part I--Description of the Test, *J. Mater.,* Vol 6 (No. 3), 1971, p 576-589
16. F. Borik and W.G. Scholz, Gouging Abrasion Test for Materials Used in Ore and Rock Crushing: Part II--Effect of Metallurgical Variables on Gouging Wear, *J. Mater.,* Vol 6 (No. 3), 1971 p 590-605
17. F. Borik, Using Tests to Define the Influence of Metallurgical Variables on Abrasion *Met. Eng. Quart.,* Vol 12 (No. 2), 1972, p 33-39
18. E.N. Gregory and M. Bartle, Materials for Hardfacing, *Weld Surfacing and Hardfacing,* The Welding Institute, Cambridge, UK, 1980, p 11-21
19. D.E. Diesburg and R. Borik, Optimizing Abrasion Resistance and Toughness in Steels and Irons for the Mining Industry, *Proceedings of Conference on Materials for the Mining Industry,* Climax Molybdenum Company, AMAX, 1974, p 15-41
20. E.N. Gregory, Selection of Materials for Hardfacing, *Weld Surfacing and Hardfacing,* The Welding Institute, Cambridge, UK, 1980, p 22-27
21. H.S. Avery, Work Hardening in Relation to Abrasion Resistance, *Proceedings of Conference on Materials for the Mining Industry,* Climax Molybdenum Company, AMAX, 1974, p 43-77
22. D.N. Rosenblatt, "The Use of High Strength Low Alloy Steels for Abrasive Applications," Technical Report C6-17.3, American Society for Metals, 1966
23. Weld Design and Fabrication, Data Sheet 420, *Weld. Des. Fabr.,* Vol 48 (No. 8), 1975, p 50
24. R.H. Sailors and J. Owens, Cast High Chromium Media in Wet Grinding, *Proceedings of Intermountain Mineral Symposium,* Climax Molybdenum Company, AMAX, 1982, p 53-62
25. T.R. Rowberry, Applications of Weld Surfacing to Power Plant, *Weld Surfacing and Hardfacing,* The Welding Institute, Cambridge, UK, 1980, p 54-57
26. T.E. Norman, A Review of Materials for Grinding Mill Liners, *Proceedings of Conference on Materials for the Mining Industry,* Climax Molybdenum Company, AMAX, 1974, p 207-218
27. H.L. Arnson, J.L. Parks, and D.R. Larsen, Alloys and Designs for Large Mill Liners and Crusher Parts, *Proceedings of Intermountain Mineral Symposium,* Climax Molybdenum Company, AMAX, 1982, p 25-42
28. H.S. Avery, Hard Surfacing by Fusion Welding, *Monogram on Wear,* American Brake Shoe Company, 1947
29. A.J. Hickl, An Alternate to Cobalt-Base Hardfacing Alloys, *J. Met.,* Vol 32 (No. 3), 1980, p 6-12
30. H.S. Avery, "Hardfacing Alloys," Technical Report C6-17.3, American Society for Metals, 1968
31. A.J. Osborn, Hardfacing in the Mining Industry, *Aust. Min.,* Vol 65 (No. 10), 1973, p 56-57, 60-61
32. G.A. Kennedy, *Welding Technology,* Howard W. Sams & Co., Inc., 1975, p 563-585
33. R.L. Chavanne, "Thirty-Five Ways to Improve Hardfacing," Rock Products, March 1976
34. R.J. Dawson, S. Shewchuck, and J.E. Pritchard, Selection and Use of Hardfacing Alloys, *Proceedings of Intermountain Mineral Symposium,* Climax Molybdenum Company, AMAX, 1982, p 109-120
35. A.M. Horsfield, Weld Surfacing Processes, *Weld Surfacing and Hardfacing,* The Welding Institute, 1980, p

32-39

36. C.J. Wilson, Evolutionary Improvement of the "Old" and the Development of New Wear Resistant Materials in Mill Liners and Cast Grinding Balls, *Proceedings of Intermountain Minerals Symposium,* Climax Molybdenum Company, AMAX, 1982, p 43-46
37. P.J. Moroz, Jr., The Effects of Matrix Hardness and Microstructure on the Wear of Steel Grinding Balls During Wet Iron Ore (Magnetite) Grinding, *Proceedings of Intermountain Minerals Symposium,* Climax Molybdenum Company, AMAX, 1982, p 63-77
38. J.C. Farge and G.A. Barclay, Properties and Performance of Cast Iron Grinding Balls, *Proceedings of Conference on Materials for the Mining Industry,* Climax Molybdenum Company, AMAX, 1974, p 189-200
39. D.E. Nass, Steel Grinding Media Used in the United States and Canada, *Proceedings of Conference on Materials for the Mining Industry,* Climax Molybdenum Company, AMAX, 1974, p 173-188
40. A.K. Gangopadhyay and J.J. Moore, Effect of Impact on the Grinding Media and Mill Liner in a Large Semiautogeneous Mill, *Wear,* Vol 114, 1987, p 249-260
41. J.J. Moore, R. Perez, A.K. Gangopadhyay, and J.F. Eggert, Factors Affecting Wear in Tumbling Mills: Influence of Composition and Microstructure, *Int. J. Min. Proc.,* Vol 22, 1988, p 313-343

Friction and Wear of Medical Implants and Prosthetic Devices

Duncan Dowson, The University of Leeds (England)

Introduction

SYNOVIAL JOINTS are remarkable bearings. In man they generally operate as dynamically loaded bearings subjected to about 10^8 cycles of loading in a 70-year lifetime. They are self-contained and subjected to the oscillatory motion associated with the requirements of human locomotion and their position in the skeletal frame.

The natural synovial joint exhibits the features of many engineering plain bearings. It consists of layers of bearing material (articular cartilage) mounted on relatively hard bones forming the skeletal frame (Fig. 1). The lubricant (synovial fluid) is a water-based fluid containing a additive (hyaluronic acid) that not only produces a tacky, viscous fluid which ensures adequate fluid-film lubrication under most conditions in healthy joints, but also combines with protein to form a very effective boundary lubricant. The slipperiness of cartilage-coated bone ends is well known, and this led to the early concept of boundary lubrication in synovial joints. However, carefully conducted experiments (Ref 1, 2) suggested that fluid-film lubrication prevailed in natural joints, and recent theoretical work (Ref 3, 4) has confirmed this possibility. It appears that natural recognized the merits of elastohydrodynamic and microelastohydrodynamic lubrication long before engineers appreciated the phenomena and adopted the to advantage in machine design.

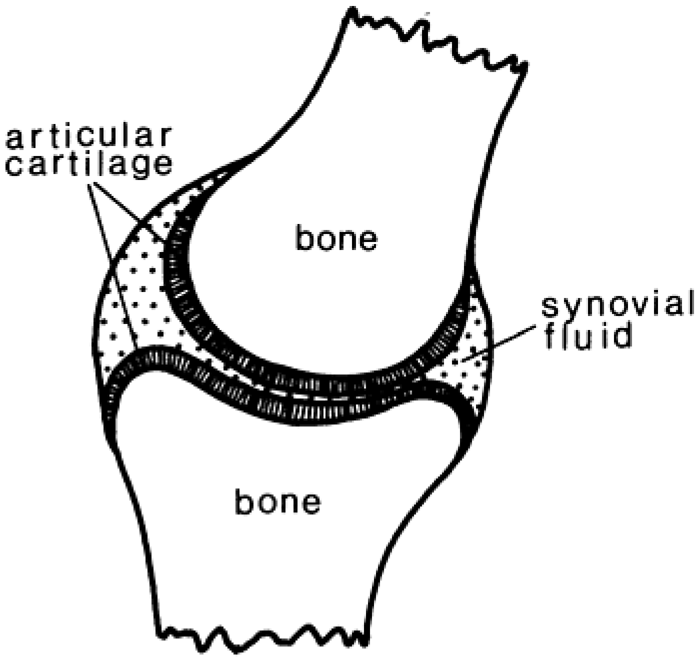


Fig. 1 Schematic showing key components of a natural synovial joint

In tribological terms, the major load-bearing natural synovial joints, such as the hip and the knee, operate with average coefficients of friction of about 0.02 and wear factors of about 10^{-6} $mm^3/N \cdot m$. Like many bearings, they tend to deteriorate with time, and in due course the bearing material roughens, fissures, and wears away to expose the underlying bone. This process, which represents but one form of the rheumatic diseases, is known as osteoarthrosis and is generally associated with a restriction or loss of movement, increased stiffness, and pain. In severe cases, surgical intervention may be necessary. It is at this stage that the materials scientist and the tribologist are called upon to work alongside the surgeon in the design and development of satisfactory total replacement joints.

Historical Background

Total joint replacement is widely regarded as the major achievement in orthopedic surgery this century. Arthroplasty, or the creation of a new joint, is the name given to the surgical treatment of degenerate joints aimed at the relief of pain and the restoration of movement. This has been achieved by excision, interposition, and replacement arthroplasty and by techniques that have developed over about 170 years.

Early Excision Arthroplasty

In excision arthroplasty, which developed in pre-anesthetic and pre-antiseptic times, bone is divided, remodeled, and removed. One of the earliest approaches to excision arthroplasty is attributed to Anthony White, who operated successfully on a 9-year-old boy at the Westminster Hospital in London in 1822 (Ref 5). Shortly afterwards, a planned excision arthroplasty for a hip deformity was reported by J.R. Barton of Philadelphia (Ref 6).

Interposition Arthroplasty

One of the main objectives of arthroplasty is to re-establish movement in the joint, but this can readily be defeated by arthrodesis--the subsequent fusion and fixation of the bones. Interposition arthroplasty attempts to overcome this problem by interposing foreign materials between the bone ends. The procedure developed in the mid-19th century, with notable contributions being reported by Gluk (Ref 7) in 1885, Jones (Ref 8) in 1902, Hey-Groves (Ref 9) in 1926, and Smith-Peterson (Ref 10, 11) between 1923 and 1938.

A remarkable range of materials was interposed between the bone ends, including muscle, wood, ivory, gold foil, celluloid, pyrex, phenol-formaldehyde resin, and a cobalt-chromium-molybdenum alloy known as Vitallium then used in

dentistry. In many of the initial cases, the insert was simply interposed between the surfaces, but in most of the latter devices the material was shaped into a cup, which allowed some movement between it and both the femoral head and the acetabulum. Interposition arthroplasty was adopted for hundreds of patients prior to the widespread introduction of total joint replacements in the 1950s and 1960s.

Replacement Arthroplasty

The origin of replacement arthroplasty is generally associated with Philip Wiles (Ref 12), who, while working at the Middlesex Hospital in 1938, designed and introduced a stainless steel total replacement hip joint. The acetabular cup and femoral head were ground to fit together accurately, and the femoral head was mounted on a bolt that passed down the neck of the femur. If appears that the six operations carried out achieved some success, but the work was disrupted during World War II and all records were lost. It nevertheless appears that this pioneering development represented the first effective introduction of metal-on-metal total replacement joints.

Femoral Head Replacements. It was not until the 1950s that further progress was reported on total replacement joints, but in the meantime a number of interesting femoral head replacements were introduced. Notable among these were the acrylic prosthesis developed by the Judet brothers in Paris in 1946 (Ref 13) and the cobalt-chromium alloy prosthesis introduced by Austin Moore in the United States in 1950 (Ref 14). The Judet prosthesis was received with some enthusiasm, but by 1954, when some 850 patients had been operated upon by the Judet brothers, there was concern about the incidence of fracture and severe abrasion of the acrylic head. Nylon was tried as an alternative material, but excessive wear took place and the wear debris was associated with acute tissue reaction. The Austin Moore implant did not suffer such wear problems, and many operations were carried out, but concern was expressed for the migration of the acetabulum and the penetration of the femoral head into the pelvis. An interesting feature of the later forms of the Austin Moore femoral implant was the fenestrations of the intramedullary stem, since it appeared that cancellous bone grew into the foramina to effect anchorage of the prosthesis.

Total Hip Replacements. Two English surgeons made outstanding contributions to the development of total hip replacements in the 1950s.

Metal-on-Metal Implants. In 1951, G.K. McKee of Norwich introduced a metal-on-metal prosthesis in which both the femoral and acetabular components were made of stainless steel (Ref 15). The acetabular cup was initially fixed into the pelvis by means of screws, but because these came loose within a year, no doubt due to the excessive friction associated with a stainless steel ball seated within a close-fitting acetabular cup, the material of construction was changed to Vitallium, a cobalt-chromium-molybdenum alloy. McKee then adopted a modified form of the Thompson femoral component developed in the United States and a cloverleaf form of acetabular cup that was again fixed by a screw. A success rate of about 50% was reported in the period of 1956 to 1960, but it became clear that success depended primarily on the metallic components remaining tight in the bone. In 1960, when McKee and his colleague Watson-Farrar introduced methyl-methacrylate as a cement to hold the components in place, the success rate rose to an encouraging 90%. By 1967, the outer surface of the cup was studded to assist fixation in the cement, as shown in Fig. 2 (Ref 16). McKee wrote confidently of the excellent chance of his total replacement hip joints lasting for 10 to 20 years. He also wrote of the "lapped-in" Vitallium components:

> ". . .the frictional resistance is small, the bearing surfaces are hard, and wear is reduced to a minimum."

McKee also recognized that the rubbing together of identical metals was tribologically unsound, but necessary in total replacement joints to avoid electrochemical corrosion:

> ". . .If metal is used in the body in the manufacture of an artificial joint, then both components must be of the same metal, otherwise an electrolytic reaction would be set up and corrosion would occur. Although this is contrary to usual engineering principles, it is essential in replacement joint surgery and it is also important to use a material, metal or other substance, which is inert even in a finely particulate form".

There has been some reawakening of interest in metal-on-metal prostheses in recent years, and it is therefore important to recognize the major contributions of McKee some 30 or 40 years ago.

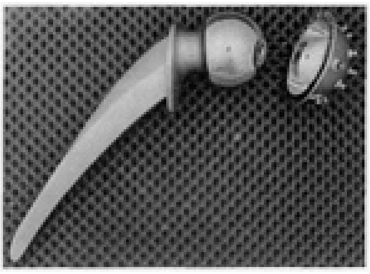


Fig. 2 McKee-Farrar total replacement hip joint incorporating lapped-in Vitallium femoral and acetabular cup components. Stud-shaped projections facilitate installation of cup into hip bone of patient.

Metal-on-Polymer Implants. The outstanding success of modern replacement joint operations is undoubtedly attributable to the late Professor Sir John Charnley. He recognized that is was necessary to combat loosening of the components by reducing the frictional torque generated by the articulating surfaces. It appears that a factor which directed him to this conclusion was the chance encounter with a patient fitted with a Judet femoral head replacement that squeaked! The outcome was his concept of a "low-friction arthroplasty" based upon a small-diameter metallic femoral head engaged in a polymeric acetabular cup. The frictional torque was thus reduced by a combination of sound biomechanics and tribology.

Charnley studied the lubrication of natural and total replacement synovial joints. He concluded that whereas synovial fluid was an exceptional lubricant for natural joints, it was not at all well suited for any of the replacement materials then considered. In 1958, he concluded that:

> ". . .the only chance of success in lubricating an artificial animal joint would be by using surfaces which were intrinsically slippery on each other; in other words, self-lubricating irrespective of whether tissue fluid were present or not."

This philosophy prompted Charnley to introduce polytetrafluoroethylene (PTFE, or Teflon) acetabular cups and stainless steel femoral stems in 1959. The early results with this remarkably low-friction polymer were spectacular, but the very poor wear characteristic of the material led to rapid penetration of the cup wall by the metallic femoral head after only 2 or 3 years (Fig. 3). Some 300 replacement hip operations were carried out between 1959 and 1961 before PTFE was abandoned.

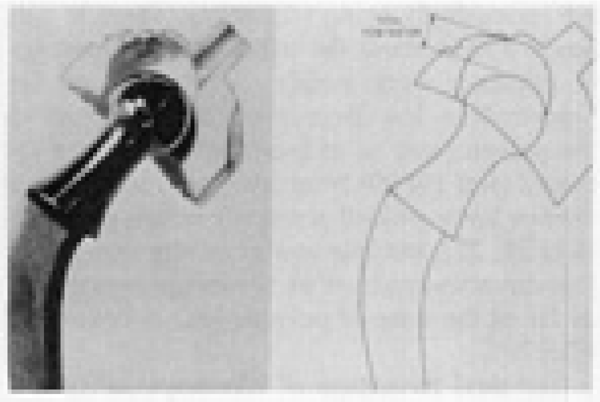


Fig. 3 Excessive penetration of PTFE acetabular cup by a stainless steel femoral stem approximately 2 to 3 years after implantation of a vintage Charnley joint in the early 1960s

Attempts were made to improve the wear resistance of the polymer by introducing fillers into the PTFE, but the improvements noted in laboratory tests were not reproduced in the body. In 1961, Charnley turned to ultrahigh molecular

weight polyethylene (UHMWPE) as the cup material. This material exhibits a higher coefficient of friction than PTFE, but a vastly enhanced resistance to wear. Its introduction was astonishingly successful, today, some 30 years later, it is still the dominant polymeric material in total replacement joints.

Alternative Metals

Strength requirements have ensured the use of metals for the femoral components of both hip and knee replacements. The range of metals used include stainless steel (such as 316L), cobalt-chromium-molybdenum alloys (such as Co-28Cr-6Mo), and titanium and titanium alloys (such as Ti-6Al-4V). Ti-6Al-4V, originally developed for the rotor and stator blades of aircraft gas turbine engines, rapidly replaced high-purity titanium, which failed to provide adequate strength for load-bearing joints (Ref 17). Stainless steel and the harder cobalt-chromium-molybdenum and titanium alloys appear to be the dominant metals in current metal-on-polymer implants.

In some cases, there has been a return to the concept of metal-on-metal prostheses, based on highly polished cobalt-chromium-molybdenum components (Ref 18). Careful design and manufacture can undoubtedly lead to very low wear rates, but there is still some concern regarding the relatively high friction and the deleterious effects of metallic wear particles.

Ceramics

High-purity aluminum oxide is potentially an attractive implant material for total replacement synovial joints, because it is bioinert and can be prepared to present an excellent counterface to polyethylene. A high-density, fine-grained Al_2O_3 is generally preferred, and hip joints in the form of polyethylene acetabular cups, ceramic femoral heads, and metallic femoral stems are increasingly being implanted. There is still much debate about the tribological advantages of ceramics over metals as counterfaces for polyethylene, but decreases in the wear rates of the polyethylene of at least 50% have been reported (Ref 19, 20) from laboratory tests. Other studies have yielded a more confusing picture (Ref 21, 22), but it is now generally recognized that ceramics can have an advantage over metals as far as the wear of polyethylene is concerned (Ref 23).

The dual advantage of biocompatibility and excellent wear resistance prompted the introduction of ceramic-on-ceramic hip replacements in the 1970s and 1980s (Ref 24, 25, 26, 27). It is necessary to produce an excellent surface finish on such implants and to hold fine tolerances on the important dimensions during precision manufacture. The unforgiving nature of the materials also calls for precise surgical techniques and careful consideration of strength requirements. Some of these implants failed due to fatigue and surface fracture, and recent clinical reports have sounded cautionary notes (Ref 28, 29).

Current Status

Total hip replacement is now a well-established orthopedic procedure, with about 400,000 hip joints being replaced by man-made alternatives each year worldwide, mainly in Europe and the United States. Knee prostheses are developing rapidly, and considerable progress will likely be reported in this field in the next decade.

The primary limitations are no longer tribological, although there is considerable evidence to suggest that tissue reaction to wear debris is associated with implant loosening, and current prostheses have achieved remarkable success as replacement rubbing bearings in the human body. Other aspects of material and implant performance, such as fixation, corrosion, mechanical strength, and surface variations, lie beyond the scope of this article and will be discussed only insofar as they affect the tribological characteristics of prostheses.

There are various types of metal porous surfaces on prostheses now. Some surfaces have additional coatings of hydroxyapatite (the mineral part of bone) to enhance initial ingrowth and to provide a more natural barrier between the bone and the implant.

Additional information can be found in the article "Corrosion of Metallic Implants and Prosthetic Devices" in *Corrosion,* Volume 13 of *ASM Handbook,* formerly 9th Edition *Metals Handbook.*

Properties of Implant Materials

It has been noted in the previous section that the most common materials used in total replacement synovial joints are metals, ceramics, and UHMWPE. The principal physical properties of these materials will be summarized before focusing on their tribological characteristics.

Metals. The introduction of chromium-containing "stainless" steels into the manufacturing industry in the early 20th century was followed by their use in orthopedic surgery from about 1926. They remain the major implant alloys in general orthopedic surgery to the present day, although specifications and manufacturing techniques have changed and they recently have been joined by the harder and more corrosion-resistant cobalt-chromium-molybdenum and titanium alloys. In the United States, 316L is not currently used very extensively for permanent implants.

Typical compositions of the implant metals used in orthopedic surgery are given in Table 1; their physical and mechanical properties are listed in Table 2. Full accounts of the biocompatibility and general physical and mechanical properties of implants metals have been presented in Ref 30, 31, 32, 33.

Table 1 Composition of implant metals and alloys used in orthopedic surgery applications

Material	Composition, wt%										
	Al	**C**	**Co**	**Cr**	**Fe**	**Mn**	**Mo**	**Ni**	**Si**	**Ti**	**V**
Type 316 stainless steel (wrought)	. . .	0.03 max	. . .	16-19	bal	2 max	2-4	10-15	1 max	. . .	. . .
Cobalt-chromium-molybdenum alloy (cast)	. . .	0.2	bal	27-30	1	1	5-7	2.5	1	. . .	. . .
Titanium (wrought)	. . .	0.5	. . .	. . .	0.2 max	. . .	. . .	. . .	. . .	bal	. . .
Titanium-aluminum-vanadium alloy	5.5-	0.08	. . .	. . .	0.3	. . .	. . .	. . .	. . .	bal	3.5-

Table 2 Physical and mechanical properties of implant metals and alloys used in orthopedic surgery applications

Material	Physical properties			Mechanical properties										
				Young's modulus		**Tensile strength**		**0.2% proof stress**		**Fracture strain, %**	**Fatigue stress, 10^8 cycles**			
	Density		**Hardness, HV**								**Air**		**Saline**	
	g/cm^3	**lb/in.^{3}**		**GPa**	**psi × 10^6**	**MPa**	**ksi**	**MPa**	**ksi**		**MPa**	**ksi**	**MPa**	**ksi**
Type 316 stainless steel (wrought)	7.90	0.285	183	200	29	465	67	170	25	40	241	35	103	15
Cobalt-chromium-molybdenum alloy (cast)	7.80	0.282	300	200	29	665	96	455	66	10	290	42	140	20
Titanium (wrought)	4.50	0.163	250	127	18.4	575	83	465	67	15	250	36	120	17
Titanium-aluminum-vanadium alloy	4.40	0.159	330	111	16.1	900	130	830	120	8	380	55	140	20

Additional information can be found in the article "Corrosion of Metallic Implants and Prosthetic Devices" in *Corrosion,* Volume 13 of *ASM Handbook,* formerly 9th Edition *Metals Handbook.*

Ceramics. High-density, fine-grained Al_2O_3 has proved to be the most widely used bioceramic in total replacements thus far. Body fluids can reduce the mechanical strength of ceramics, primarily through the ingress of liquid into the pores of low-density material; therefore, it is necessary to use high-density materials in orthopedic implants. Listed below are representative properties of alumina ceramics used in implants:

Chemical composition, %	>99.5 Al_2O_3
Density, g/cm^3 (lb/in.^{3})	>3.9 (>0.14)
Average grain size, μm	<7
Microhardness, GPa (kgf/mm^2)	23 (2.3 × 10^3)
Compressive strength, MPa (kgf/mm^2)	>4 × 10^3 (>400)
Flexural strength, MPa (kgf/mm^2)	>400 (>40)
Young's modulus, GPa (psi × 10^6)	38 (5.5)
Impact strength, MPa (kgf/mm^2)	>0.40 (>0.04)
Corrosion resistance (Ringer's solution), mg/m^2 per day	<0.1

Ultrahigh Molecular Weight Polyethylene. It is worth recalling that it was only in 1939 that ICI Ltd. produced low-density (specific gravity, 0.91 to 0.935) polyethylene (LDPE) by the polymerization of ethylene gas and that high-density (specific gravity, 0.94 to 0.97) polyethylene (HDPE) was first produced in Europe and the United States in the 1950s. Thus, the introduction of UHMWPE into orthopedic implants early in the 1960s represented a very rapid application of a new biomaterial, which necessarily meant that many of its tribological characteristics under physiological conditions were imperfectly understood. Indeed, a comprehensive picture of the biotribological features of this remarkable material is only now emerging (Ref 34). The tensile and physical properties of UHMWPE are listed in Table 3.

Table 3 Tensile properties of RCH 1000 surgical-grade UHMWPE plastic at selected strain rates

RCH 1000 has the following physical properties: molecular weight, 3.5 × 10^6 to 4.0 × 10^6; density, 0.94 g/cm^3 (0.034 lb/in.^{3}); hardness (at 20 °C, or 70 °F), 5.3 ± 0.3 HV; bulk shear stress, 24 MPa (3.5 ksi).

Tensile property	Strain rate	
	3.28 × 10^{-3} s^{-1}	**3.28 × 10^{-2} s^{-1}**
Gage length, mm (in.)	25.4 (1.0)	25.4 (1.0)
Yield stress, MPa (ksi)	20.3 (2.94)	24.9 (3.61)
Fracture stress, MPa (ksi)	28.4 (4.12)	30.6 (4.44)

Tribological Characteristics of UHMWPE

The wear characteristics of UHMWPE have been studied extensively ever since the material was first introduced into orthopedic implants some 30 years ago. There have been three major types of investigations during this period:

- Studies of the basic wear mechanisms on standard laboratory machines (for example, pin-on-disk and pin-on-plate configurations)
- Evaluations of complete implants in joint simulators
- *In vivo* assessment of the performance of total replacement joints

Pin-on-Disk Experiments

The general form of a tri-pin-on-disk friction and wear measuring machine is shown in Fig. 4. One of the test materials, usually the metal or ceramic, is machined in the form of a disk and located on pegs on a turntable driven at constant speed in an environmental chamber. Three test pins, usually the polymer, are mounted in a frame and loaded against the turntable by means of static weights or a hydraulic system. The frame is prevented from rotating by means of three cantilever arms, whose deflection indicates the friction torque on the frame.

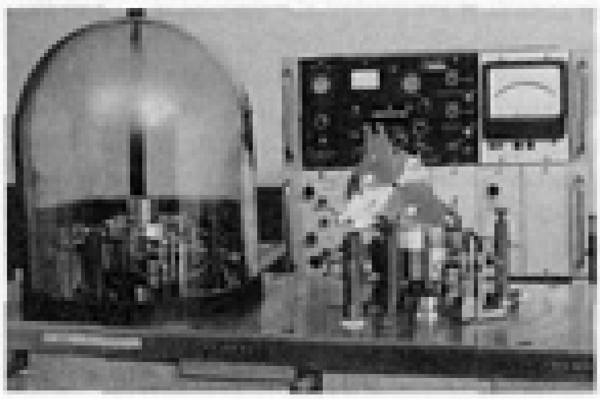


Fig. 4 Typical tri-pin-on-disk friction and wear testing machine that measures wear under stable speed, load, and environmental conditions

The wear pins are generally cylindrical in form, with flat, spherical, or truncated conical ends. The latter configuration, shown in Fig. 5, has the advantage of offering a substantial foundation for the wear face, while permitting wear to take place over a relatively small area.

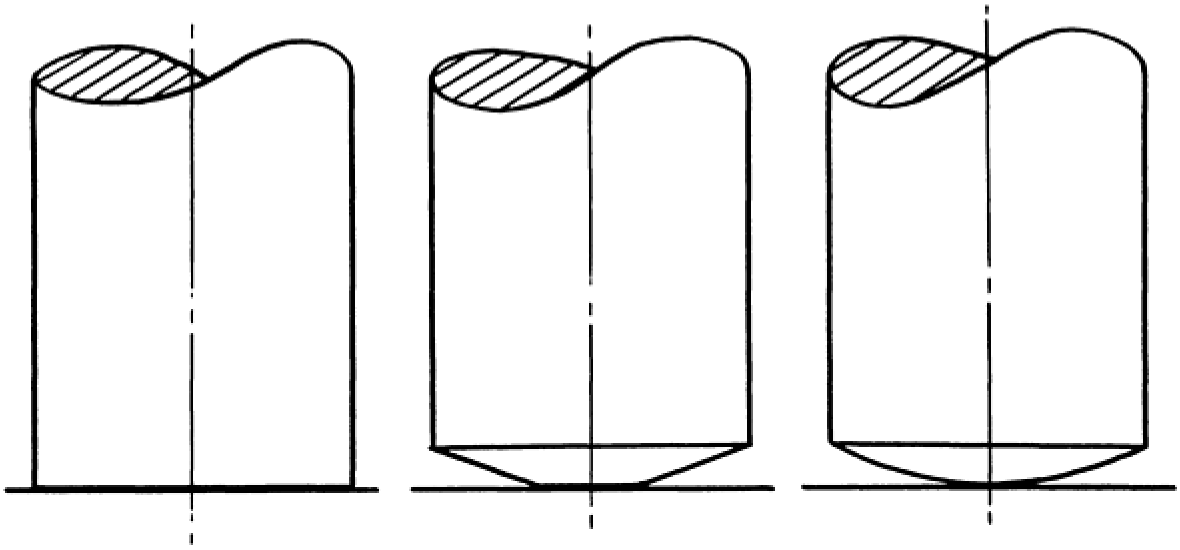


Fig. 5 Schematic showing typical geometry of wear pins evaluated on a tri-pin-on-disk friction and wear testing machine. (a) Flat end. (b) Truncated conical end. (c) Spherical cap end

Wear can be measured by monitoring dimensional changes, such as the diameter of the wear face of the length of the pin, or by measuring weight loss. The former method has much appeal, but with polymers the results generally exhibit considerable scatter. The reason for this is that the dimensions being measured are influenced by both temperature and humidity. Therefore, the measurement of weight loss is generally preferred, although even with this method considerable care has to be exercised to measure small changes in the liquid content of the polymer specimens associated with the lubricant or the humidity of the atmosphere.

A great advantage of the pin-on-disk machine is that it enables wear to be studied under stable conditions of speed, load, and environment. If the weight or volume changes associated with wear are monitored periodically and plotted against sliding distance, a linear relationship is generally observed. This enables the wear factor (k) to be ascertained from the simple relationship between wear volume (V) and the load (P) and sliding distance (X):

$$V = kPX \qquad \textbf{(Eq 1)}$$

The wear characteristics of different combinations of materials can thus be quantified and compared using the wear factor, *k*, enabling initial screening of potential sliding pairs of materials to be made for a particular application. Furthermore, the transfer film that is formed by the polymer on the hard counterface under many conditions can be studied throughout the test, and the surfaces of both the monolithic polymeric specimens and the transfer film can be subjected to surface analysis and examination by electron microscopy after testing.

A valuable and early summary of wear test methods and a compilation of wear rates and factors from many source has been presented by Dumbleton (Ref 35). It is generally recognized that short-term tests can be misleading, since the generation of a reasonably steady form of transfer film takes a finite time. Furthermore, there is evidence that surface fatigue plays a major role in determining the long-term wear rate of polyethylene sliding against steel (Ref 36, 37), and this characteristic might be totally missed in short-term tests. Dowson (Ref 37) has reported increases in the wear factor of about 60 to 70% following the onset of surface fatigue in the polymer in steady-state sliding wear in a tri-pin-on-disk machine under dry (atmospheric) conditions.

Wear factors for UHMWPE sliding on very smooth (R_a = 0.013 to 0.036 μm) EN58J stainless steel disks were found to be of the order of 10^{-7} mm^3/N · m under dry conditions and 10^{-9} to 10^{-8} mm^3/N · m in the presence of distilled water. It should be noted that the counterface roughness was exceptionally good, and certainly better than that achieved in most metallic implants. It was also reported that when UHMWPE slid on itself under dry conditions, the wear factor was much higher (about 10^{-4} mm^3/N · m).

Perhaps the most important feature of these early wear factors obtained for UHMWPE was their exceptionally low values, indicating very good resistance to wear. Indeed, the values were so low that they were inconsistent with the *in vivo* performance of total replacement hip joints, thus casting doubts on the value of pin-on-disk measurements of wear as far as implant performance was concerned. The measurement of the very small wear rates of a UHMWPE was subsequently discussed further by McKellop *et al.* (Ref 38), Clarke (Ref 39), and Dumbleton (Ref 35). An equally important observation was the evidence of surface fatigue in the polymer after a significant sliding distance (Ref 36).

Pin-on-Plate (Reciprocating) Experiments

Human joints experience oscillatory sliding conditions, and thus much research has been carried out on reciprocating pin-on-plate, rather than steady-speed pin-on-disk, machines. Such testing still does not simulate the exact motion encountered in synovial joints, but it does include the essential feature reversals in motion and hence in surface shear stresses. A six-station pin-on-plate friction and wear machine is shown in Fig. 6.

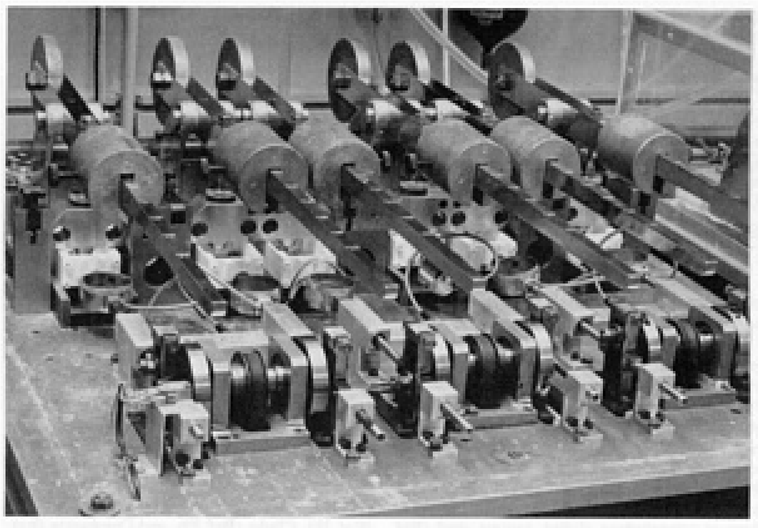


Fig. 6 Six-station reciprocating pin-on-plate friction and wear machine used to evaluate prosthetic implant materials

This type of testing machine is very useful for the initial screening of many potential implant materials. It has also proved to be particularly valuable in revealing important aspects of the role of the counterface roughness and the environment. Typical sets of data for the deduced wear factors for UHMWPE pins loaded against EN58J stainless steel in both dry (atmospheric) and wet (distilled water) environments are shown as functions of counterface roughness in Fig. 7. It is clear the long-term wear factor exhibits a minimum at a counterface roughness slightly less than 0.1 μm. Furthermore, the presence of water increases the wear factor substantially, typically by an order of magnitude, for counterface roughnesses representative of many implants (R_a values of 0.05 to 0.1 μm). Because replacement joints appear to enjoy a measure of lubrication, it thus appears from these pin-on-plate reciprocating tests that wear factors on the order of 10^{-6} m^3/N · m can be expected in implants.

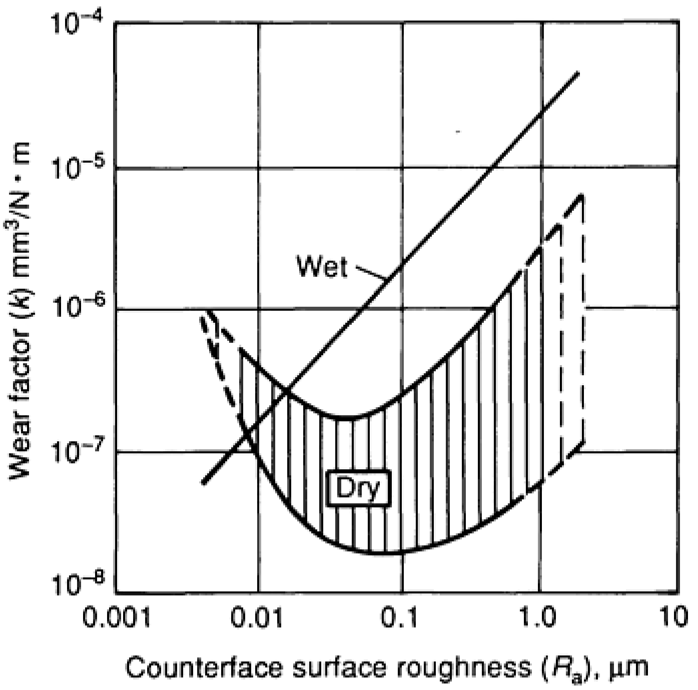


Fig. 7 Plot of UHMWPE plastic wear factor versus stainless steel counterface surface roughness evaluated under both dry (atmospheric) and wet (distilled water) conditions

Joint Simulators

It is customary to evaluate the performance of engineering tribological components in machine simulators prior to their introduction into products, yet the use of joint simulators is a relatively recent development. There are still few satisfactory joint simulators in service, but it is increasingly important that new biomaterials and new implant designs be fully tested in the laboratory before their release for clinical evaluation.

Hip Joint Simulators. Modern forms of hip joint simulators, such as that shown in Fig. 8, subject the implant to the dynamic loads and the complex motions experienced in walking or some other simple, specified motion. The computer-controlled load components are usually applied hydraulically, in mutually perpendicular directions, with a feedback mechanism to ensure that the specified load is indeed applied. Joint simulators need to be robust; otherwise, the long-duration tests become critical evaluations of the seals, bearings, and hydraulic and mechanical components of the testing machine, rather than of the total replacement joints themselves.

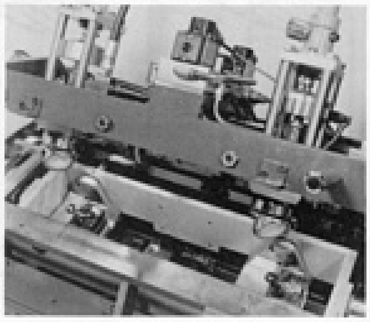


Fig. 8 Multistation hip joint simulator duplicates walking and other human motions typically encountered by prosthetic implants

The evaluation of wear in implants tested in joint simulators is more difficult than in the simple forms of laboratory friction and wear testing machines described earlier, because the component being evaluated, the acetabular cup, is a good deal heavier than the simple wear pins used in pin-on-disk or pin-on-plate machines. This use of heavy components makes it difficult to achieved adequate accuracy in the determination of small weight changes. Furthermore, the acetabular cups are usually embedded in a heavy metallic holder, probably with acrylic cement, in order to replicate as closely as possible the *in vivo* situation.

Dimensional changes thus feature much more prominently than weight loss determination in the evaluation of penetrations of femoral heads into acetabular cups in joint simulators. Casting and shadowgraph techniques have been described in Ref 40, while coordinate measuring machines (Ref 41) have been shown to offer great advantages. It should also be noted that dimensional changes reveal the total penetration of the femoral head into the acetabular cup, and that this represents a combination of wear and creep. Some simulators (Ref 40) thus include a separate creep station, in which the dynamic loading cycles are applied to the implant without articulation. The resulting creep penetration can then be deducted from the total penetration attributable to wear plus creep in order to establish the relative contributions of each process to the overall penetration. Total penetration, of course, is of primary concern to the patient and the surgeon, but the separate contributions of wear and creep are of interest to the tribologist and the materials scientist.

An early account of the use of a hip joint simulator was reported by Duff-Barclay and Spillman (Ref 42), while Dumbleton provided an excellent account of the state of the art in the early 1980s (Ref 35). Further accounts of joint simulator studies can be found in Ref 43, 44, 45, 46.

Knee Joint Simulators. Relatively little work has been done on the development of knee joint simulators, but accounts of the principles involved can be found in Ref 35 and 47. A typical knee joint simulator, designed to operate on principles similar to those outlined for the hip joint simulator, is shown in Fig. 9. Detailed accounts of the application of holographic measuring techniques in the evaluation of the magnitude and distribution of wear in various forms of total replacement knee joints tested in the simulator can be found in Ref 48 and 49.

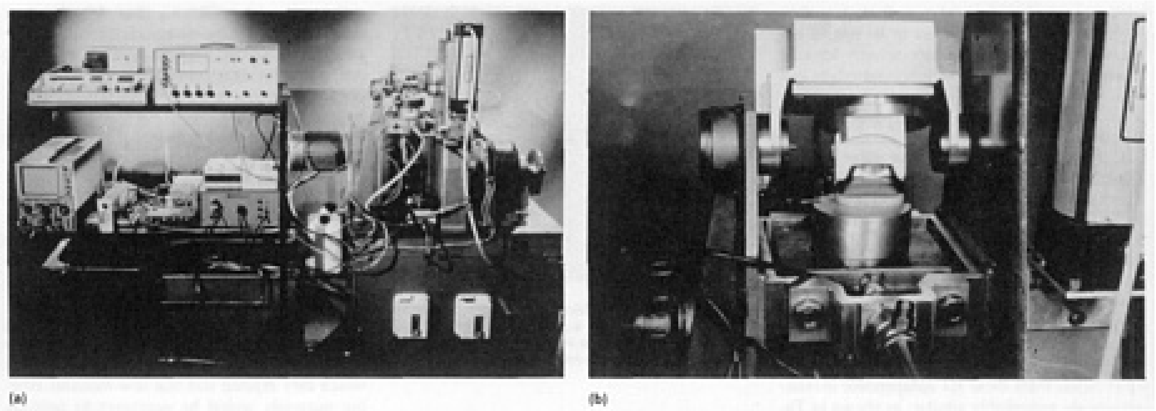


Fig. 9 Knee joint simulator used to evaluate magnitude and distribution of wear in a total replacement knee joint. (a) Complete machine setup showing test apparatus and accompanying state-of-the-art monitoring/recording instrumentation. (b) Closeup view of test cell showing knee joint implant being evaluated while it is held in position in a test fixture

In Vivo Assessment of Total Joint Replacement Performance

The penetration of femoral heads into the acetabular cups of total replacement hip joints can be assessed by radiographic studies *in vivo* or by direct measurement of explanted prostheses at the time of revision surgery or postmortem. Charnley (Ref 50) and his colleagues pioneered the former approach, and with Halley in 1975 (Ref 51) he found an average rate of penetration of 0.15 mm/year for 72 patients operated on in 1962 and 1963. Griffith *et al.* (Ref 52) subsequently assessed 491 hips from patients operated on between 1967 and 1968, and in a much-quoted paper they recorded an average rate of penetration of only 0.07 mm/year. The mean follow-up period was 8.3 years (a range of 7 to 9 years), while a few (4%) of the sockets exhibited penetration rates as high as 0.24 mm/year. An interesting feature of the measurements was that the rate of penetration generally decreased with time in the body.

More recent studies of the rate of penetration of femoral heads into UHMWPE acetabular cups in Charnley prostheses (Ref 53, 54) have confirmed that average values range from 0.07 to 0.15 mm/year. Implants removed during revision operations appear to exhibit rather larger penetration rates, typically 0.2 mm/year, as reported by Atkinson *et al.* (Ref 55). This increase may well reflect inadequate biomaterial, surgical, or mechanical factors associated with the implant and be one of the prime indicators for revision surgery.

Dumbleton (Ref 35) reviewed radiographic measurements of penetration, and in recent years many reports have been made for different forms of prostheses. Useful records are provided in Ref 56 and 57.

Dowson and Wallbridge (Ref 58) compared their laboratory wear data with a retrospective assessment of the clinical performance of early Charnley prostheses based on PTFE acetabular cups. They adapted the simple wear equation (Eq 1) to account for creep and the dynamic loading and motion cycle encountered in the hip joint:

$$k_{\text{clinical}} = \left[\pi r^2 P\left(1 - \frac{C}{100}\right)\right] \Big/ N \int L \cdot dx \qquad \textbf{(Eq 2)}$$

where P is the annual linear penetration rate in millimeters per year, r is the radius of the femoral head in millimeters, C is the percentage of the penetration attributable to creep, N is the number of loading cycles per annum, L is the instantaneous load in newtons at any point in the cycle during which the corresponding sliding distance (dx) is in meters. For the loading cycle revealed by Paul (Ref 59) for level walking and a 22 mm diam Charnley prosthesis, the denominator on the right-hand side of Eq 2 can be written as:

$$N \int L \cdot dx = 1.188\, NBd \qquad \textbf{(Eq 3)}$$

where the units are N · m/year and where B is the body weight in newtons and d is the diameter of the femoral head.

An important unknown variable in such analyses is patient activity, but on the assumption that the number of loading cycles on a 60-year-old patient fitted with a replacement hip joint would be about 1.5 million per year, the authors found that the clinical wear factor equivalent to the recorded average penetration rate for the early Charnley implants of 2.26 mm/year was 3.7×10^{-5} $mm^3/N \cdot m$. This compared favorably with the average wear factor for PTFE sliding on stainless steel of 3.4×10^{-5} $mm^3/N \cdot m$ in pin-on-disk experiments. This comparison was made on the basis of zero creep, which is now known to be incorrect, and rather thin data on patient activity, but it provided welcome support for the view that clinical performance of implants could indeed be predicted from laboratory wear data.

An assessment of the walking activity of both normal subjects and patients fitted with total replacement hip joints has been provided by Wallbridge and Dowson (Ref 60). They found that the number of loading cycles applied to each hip joint per annum (N) was related to age (A) in years by the approximate relationship:

$$N = (3.29 - 0.032A) \times 10^6 \quad \textbf{(Eq 4)}$$

When Atkinson *et al.* (Ref 55) attempted to repeat the exercise for the later range of UHMWPE Charnley implants, the agreement between laboratory and clinical performance was somewhat less satisfactory. The average clinical wear factor derived from measurements on 25 explanted prostheses was 2.9×10^{-6} $mm^3/N \cdot m$, which the results ranging from 0.9×10^{-6} to 7.2×10^{-6} $mm^3/N \cdot m$. The average surface roughness was found to be 0.054 μm, and the corresponding laboratory wear factor for this roughness was 1.2×10^{-6} $mm^3/N \cdot m$, a mere 41% of the average deduced clinical wear factor.

It was at this stage that the role of surface blemishes in the form of fine scratches on the hard metallic counterface upon the wear of the polyethylene was recognized. It had previously been shown (Ref 61) that for isotropic counterface roughness, the relationship between the wear factor (k) and the roughness average (R_a) of the hard counterface was given approximately by:

$$k = (4.0 \times 10^{-5})\, R_a^{1.2}\ mm^3/N \cdot m \quad \textbf{(Eq 5)}$$

This was the relationship used to indicate the extent of the agreement between the clinical and laboratory wear factors, but the disproportionate influence of a single blemish was revealed in Ref 62. It was shown that the piled-up edge material associated with a single transverse scratch on the stainless steel counterface could increase the wear rate by about one order of magnitude, which immediately explained why scratched femoral heads could penetrate the acetabular cups at a greatly enhanced rate. Barium sulfate in the acrylic cement appears to be one of the abrasive substances that may scratch femoral heads. The selection of the hard counterface in a total joint replacement must thus represent a balance between hardness and a good surface finish.

Ceramic femoral heads have attracted considerable attention in recent years. They exhibit excellent biocompatibility and good bearing surfaces for rubbing conditions. The wear rate of UHMWPE against well-polished Al_2O_3 does indeed appear to be lower than that against the metals currently employed in orthopedic implants, not only in the laboratory (Ref 19, 20) but also *in vivo* (Ref 63, 64, 65, 66, 67, 68). The ratio of the rates of penetration of the femoral heads into the UHMWPE acetabular cups for ceramic and metallic heads from these six independent investigations are remarkably similar, as shown in Table 4, and totally consistent with the laboratory studies reported in Ref 19 and 20. The indications that ceramic femoral heads will take roughly twice as long as metallic femoral heads to penetrate to a given extent into UHMWPE acetabular sockets are now firmly established. An example of a Charnley low-friction prosthesis with either metallic (stainless steel) or ceramic (Al_2O_3) femoral heads and a UHMWPE acetabular cup is shown in Fig. 10.

Table 4 Ratio of penetration rates of ceramic and metallic femoral heads into the UHMWPE acetabular cup component of Charnley prostheses

Penetration rate of femoral head material into acetabular socket		Ratio of penetration rates of femoral heads into acetabular cups[a]	Ref
Ceramic, mm/year	**Metallic, mm/year**		
. . .	. . .	0.56	63
0.23	0.30 (0.31)	0.77 (0.74)	64
0.08	0.20	0.40	65
0.098	0.247	0.40	66
0.025	0.043	0.58	67
0.08	0.14	0.57	68

(a) Penetration rate ratio = Ceramic penetration rate/Metallic penetration rate

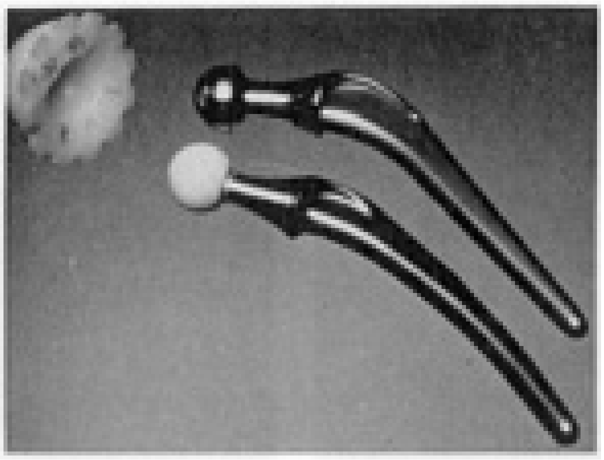


Fig. 10 Charnley low-friction hip joint prosthesis equipped with a polymeric UHMWPE acetabular cup and either a stainless steel (top) or an Al_2O_3 ceramic (bottom) femoral head

Friction. The wear of acetabular cups and the resulting penetration of both metallic and ceramic femoral heads into the sockets in the pelvis have been emphasized thus far. Brief reference should also be made to friction. The frictional torque exerted on both natural and total replacement joints has been studied in pendulum machines and joint simulators.

Natural synovial joints enjoy a mixture of fluid-film and mixed lubrication (Ref 69), but a representative coefficient of friction is 0.02. All current forms of total replacement joints rely on boundary or at best mixed-film lubrication, and their coefficients of friction are therefore inevitably higher than those of their natural counterparts. Some care has to be exercised in quoting coefficients of friction, because such values are usually derived from torque measurements and normalized in terms of the radius of the femoral head. Fricker (Ref 70) has presented a clear statement of the problem. Dumbleton has reviewed measurements of coefficients of friction for prosthetic materials (Ref 35) and shown that the value for UHMWPE and various counterfaces range from 0.02 to 0.25, with a typical value being about 0.07. Detailed investigations of the friction of prostheses have also been reported in Ref 71, 72, 73. Friction in itself does not appear to limit the succesful operation of total joint replacements, but it can be a major factor in loosening of the implant and subsequent loss of function.

Alternative Materials and Future Prospects

Because penetration rates of the femoral head into the acetabular cup in current forms of total replacement joints are typically in the range of 0.1 to 0.2 mm/year, and overall migrations of the head of 2 or 3 mm can generally be accommodated in current designs, the wear lives of implants should be in the range of 10 to 30 years. There is growing evidence to support this prediction, which means that many elderly patients can receive joint replacements with much relief and confidence. The primary problem is no longer wear, but joint loosening. However, there is increasing concern

about the role of wear debris in promoting implant loosening, and thus tribologists are being called upon to reduce the volume of wear debris substantially.

The metals, ceramics, and UHMWPE currently used in prostheses have established a sound basis for satisfactory implant performance. Indeed, it is unlikely that major advances can be expected in the reliable tribological performance of rubbing bearings for the body. The basic wear mechanisms in total replacement joints are now well understood. If the counterfaces are rough (Ref 61) or damaged (Ref 62), the basic abrasive and adhesive wear mechanisms dominate, but as the counterface roughness is improved, subsurface fatigue plays an increasingly important role. This has been clearly revealed by birefringent techniques (Ref 74) applied to implant wear by procedures originally outlined by Gibbons (Ref 75). There is much interest in the prospects for alternative structural materials, such as composites (Ref 76) and isoelastic arrangements designed to have properties similar to the bone into which they are implanted (Ref 77).

If substantial improvements are to be made in the tribological performance of synovial joints, it seems likely that new concepts will be necessary, rather than incremental improvements in rubbing bearings. One possibility that is attracting some interest would be to design an implant that would generate a fluid film between the bearing surfaces and thus emulate nature in promoting fluid-film lubrication (Ref 69). It has been pointed out that the mechanical properties of current forms of implant materials are quite different from those of the natural materials which they replace and that low-modulus bearing materials would be necessary to promote fluid-film lubrication. This concept is known as the "cushion bearing" (Ref 78).

References

1. A. Unsworth, D. Dowson, and V. Wright, The Frictional Behaviour of Human Synovial Joints, Part I. Natural Joints, *J. Lubr. Technol. (Trans. ASME),* Vol 97 (No. 3), 1975, p 369-376
2. J. O'Kelly, A. Unsworth, D. Dowson, D.A. Hall, and V. Wright, A Study of the Role of Synovial Fluid and Its Constituents in the Friction and Lubrication of Human Hip Joints, *Eng. Med.,* Vol 7 (No. 2), 1978, p 73-83
3. D. Dowson and Z.-M. Jin, Micro-Elastohydrodynamic Lubrication of Synovial Joints, *Eng. Med.,* Vol 5 (No. 2), 1986, p 63-65
4. D. Dowson and Z.-M. Jin, Micro-Elastohydrodynamic Lubrication of Low-Elastic-Modulus Solids on Rigid Substrates, *J. Phys. D, Appl. Phys.,* Vol 25 (No. 1A), 1992, p A116-A123
5. R. Mercer, *Anthology of Orthopaedics,* E & S Livingstone, London, 1966
6. J.R. Barton, On the Treatment of Ankylosis by the Formation of Artificial Joints, *N. Am. Med. Surg. J.,* Vol 3, 1827, p 29
7. T. Gluk, Referat über die durch das moderne chirurgische Experiment gewonnenen positiven Resultate, betreffend die Naht und den Ersatz von Defecten höherer Gewebe, sowie über die Verwerthung resorbirbarer und lebendiger Tampons in der Chirurgie, *Arch. Klin. Chir.,* Vol 41, 1891, p 186
8. R. Jones, On the Production of Pseudoarthrosis of the Hip Without Disarticulation of the Head, *Br. Med. J.,* Vol 1, 1908, p 494
9. E.W. Hey-Groves, Some Contributions to the Reconstructive Surgery of the Hip, *Br. Med. J.,* Vol 14, 1926-1927, p 486
10. M.N. Smith-Peterson, Joint Ankylosis: Surgical Measures for Its Prevention and Relief, *Transactions of the Interstate Post-Graduate Medical Assembly of North America,* 1925
11. M.N. Smith-Peterson, Arthroplasty of the Hip: A New Method, *J. Bone Joint Surg.,* Vol 21 (No. 2), 1939, p 269
12. P. Wiles, The Surgery of the Osteo-Arthritic Hip, *Br. J. Surg.,* Vol 45, 1957, p 488
13. J. Judet and R. Judet, The Use of an Artificial Femoral Head for Arthroplasty of the Hip Joint, *J. Bone Joint Surg.,* Vol 32B, 1950, p 166
14. A.T. Moore, The Moore Self-Locking Vitallium Prosthesis in Fresh Femoral Neck Fractures. A New Low Posterior Approach, *Am. Acad. Orthoped. Surg.,* Vol 16, 1959, p 309
15. G.K. McKee and J. Watson-Farrar, Replacement of Arthritis Hips by the McKee-Farrar Prosthesis, *J. Bone Joint Surg.,* Vol 48B (No. 2), 1966, p 245

16. G.K. McKee, Developments in Total Hip Replacement, *Proc. Inst. Mech. Eng.,* Vol 181 (Part 3J), 1967, p 85-89
17. H.S. Dobbs and J.T. Scales, Behaviour of Commercially Pure Titanium and Ti-318 (Ti-6Al-4V) in Orthopaedic Implants, *Titanium Alloys in Orthopaedic Implants,* STP 796, H.A. Luckey and F. Kubli, Jr., Ed., ASTM, 1983, p 174
18. M. Semlitsch, Twenty Years of Sulzer Experience With Artificial Hip Joint Materials, *J. Eng. Med.,* Vol 203H, 1989, p 159
19. D. Dowson and I.W. Linnett, A Study of the Wear of Ultra High Molecular Weight Polyethylene Against a High Alumina Ceramic, *Advances in Biomaterials,* Vol 2, *Mechanical Properties of Biomaterials,* G.W. Hastings and D.F. Williams, Ed., John Wiley & Sons, 1980, p 3-26
20. D. Dowson and R.T. Harding, The Wear Characteristics of Ultra-High Molecular Weight Polyethylene Against a High Density Alumina Ceramic Under Wet (Distilled Water) and Dry Conditions, *Wear,* Vol 75, 1982, p 313-331
21. M. Semlitsch, M. Lehman, H. Weber, E. Dörre, and H. Willert, New Prospects for a Prolonged Functional Life-Span of Artificial Hip Joints by Using the Material Combination of Polyethylene/Aluminium Oxide Ceramic/Metal, *J. Biomed. Mater. Res.,* Vol 11, 1977, p 537
22. I.C. Clarke, W. Phillips, H.A. McKellop, J. Moreland, and H.C. Amstutz, Sialon Ceramic--A Candidate Material for Total Joint Replacements, *Advances in Biomaterials,* Vol 2, *Mechanical Properties of Biomaterials,* G.W. Hastings and D.F. Williams, Ed., John Wiley & Sons, 1980, p 155
23. H. Oonshi, H. Igaki, and Y. Takayama, Comparisons of Wear of UHMW Polyethylene Sliding Against Metal and Alumina in Total Hip Prostheses, *Bioceramics,* Vol 1, H. Oonshi, A. Aoki, and K. Sawai, Ed., Ishiyaku Euro America, Tokyo, 1989, p 272
24. P. Boutin, Arthroplastie Totale de la hanche par Prothese en Alumine Fritee, *Rev. Chir. Orthoped.,* Vol 58, 1972, p 229
25. P. Griss, G. Heimke, H. von Andrian-Werburg, B. Krempien, S. Reipa, H.J. Lauterbach, and H.J. Hartung, Morphological and Biomechanical Aspects of Al_2O_3 Ceramic Joint Replacement. Experimental Results and Design Considerations for Human Endoprostheses, *J. Biomed. Mater. Res. Symp.,* Vol 6, 1975, p 177
26. P. Boutin, P. Christel, J.M. Dorlot, A. Meunier, A. Roquancourt, S. Hermen, L. Sedel, and J. Witvoet, The Use of Dense Alumina-Alumina Ceramic Combination in Total Hip Replacement, *J. Biomed. Mater. Res.,* Vol 22, 1988, p 1203
27. H. Mittelmeier, Anchoring Hip Endoprosthesis Without Bone Cement, *Engineering in Medicine,* Vol 2, *Advances in Artificial Hip and Knee Joint Technology,* M Schaldach and D. Hohmann, Ed., Springer-Verlag, Berlin, 1976, p 387
28. A. Walter and W. Plitz, Wear Mechanism of Alumina-Ceramic Bearing Surfaces of Hip-Joint Prostheses, *Biomaterials and Biomechanics,* Elsevier, Amsterdam, 1984, p 55-60
29. O.M. Mahoney and J.H. Dimon, Unsatisfactory Results With a Ceramic Total Hip Prosthesis, *J. Bone Joint Surg.,* Vol 72A, 1990, p 663
30. J. BuPark, *Biomaterials--An Introduction,* Plenum Press, 1979, p 1-251
31. J. Black, *Orthopaedic Biomaterials in Research and Practices,* Churchill Livingstone, 1988, p 1-394
32. D.F. Williams, Ed., *Biocompatibility of Implant Materials,* Sector Publishing, London, 1976, p 1-220
33. G.W. Hastings and D.F. Williams, Ed., *Advances in Biomaterials,* Vol 2, *Mechanical Properties of Biomaterials,* John Wiley & Sons, 1980, p 1-566
34. H.G. Willert, G.H. Buchhorn, and P. Eyerer, Ed., *Ultra-High Molecular Weight Polyethylene as Bio Material in Orthopaedic Surgery,* Hogrefe & Huber, Göttingen, 1991, p 1-272
35. J.H. Dumbleton, *Tribology of Natural and Artificial Joints,* Elsevier, Amsterdam, 1981, p 1-460
36. K.J. Brown, J.R. Atkinson, and D. Dowson, The Wear of Ultra-High Molecular Weight Polyethylene and a Preliminary Study of Its Relation to the In-Vivo Behaviour of Replacement Hip Joints, *Wear,* Vol 40, 1976, p 225
37. D. Dowson, Tribological Characteristics of Polymers With Particular Reference to Polyethylene, *Polymer Surfaces,* D.T. Clark and W.J. Feast, Ed., John Wiley & Sons, 1978, p 399

38. H. McKellop, I.C. Clarke, K.L. Markolf, and H.C. Amstutz, Wear Characteristics of UHMW Polyethylene: A Method for Accurately Measuring Extremely Low Wear Rates, *J. Biomed. Mater. Res.,* Vol 12, 1978, p 895-927
39. I.C. Clarke, Wear of Artificial Joint Materials. 1. Friction and Wear Studies: Validity of Wear-Screening Protocols, *Eng. Med.,* Vol 10 (No. 3), 1981, p 115-122
40. D. Dowson and B. Jobbins, Design and Development of a Versatile Hip Joint Simulator and a Preliminary Assessment of Wear and Creep in Charnley Total Replacement Hip Joints, *Eng. Med.,* Vol 17 (No. 3), 1988, p 111-117
41. S.N. Collins, "Evaluation of the Mechanical and Tribological Performance of Total Replacement Hip Joints in a Hip Joint Simulator," Ph.D. thesis, The University of Leeds, 1991
42. I. Duff-Barclay and D.T. Spillman, Total Human Hip Joint Prothesis--A Laboratory Study of Friction and Wear, *Lubrication and Wear in Living and Artificial Human Joints,* Institution of Mechanical Engineers, London, 1967, p 90-103
43. P.G. Niederer, M. Semlitsch, E. Dörre, and C. Dietsch, Total Hip Arthroplasty With Ceramic-Polyethylene Articulation, *Proceedings of XIV SICOT World Congress* (Kyoto), 1978, p 15-20
44. K.W.J. Wright and J.T. Scales, The Use of Hip Simulators for the Evaluation of Wear of Total Hip Prostheses, *Evaluation of Biomaterials,* G.D. Winter, J.L. Leray, and K. de Groot, Ed., John Wiley & Sons, 1980
45. J.A. Davidson and G. Schwartz, Wear, Creep, and Frictional Heat of Femoral Implant Articulating Surfaces and the Effect of Long-Term Performance. Part I. A Review, *J. Biomed. Mater. Res.,* Vol 21 (No. A3), 1987, p 261-285
46. P. Eyerer, M. Kurth, H.A. McKellup, and H. Mittlemeier, Characterization of UHMWPE Hip Cups Run on Joint Simulators, *J. Biomed. Mater. Res.,* Vol 21, 1987, p 275-291
47. D. Dowson, B. Jobbins, J. O'Kelly, and V. Wright, A Knee Joint Simulator, *Evaluation of Artificial Joints,* D. Downson and V. Wright, Ed., Biological Engineering Society, 1977, p 79-90
48. D. Downson, B.J. Gillis, and J.R. Atkinson, Penetration of Metallic Femoral Components Into the Polymeric Tibial Components Observed in a Knee Joint Simulator, *Polymer Wear and Its Control,* L.-H. Lee, Ed., Symp. Series 287, American Chemical Society, 1985, p 215-228
49. D. Dowson, P.J.J. McCullagh, and V. Wright, An Assessment of the Relative Importance of Wear and Creep in the Overall Performance of Load-Bearing Total Replacement Knee Joints, *Ultra-High Molecular Weight Polyethylene as Biomaterial in Orthopaedic Surgery,* H.-G. Willert, G.H. Buchhorn, and P. Eyerer, Ed., Hogrefe & Huber, Göttingen, 1991, p 32-40
50. J. Charnley, *Low Friction Arthroplasty of the Hip,* Springer-Verlag, Berlin, 1979, p 1-376
51. D.K. Halley and J. Charnley, Rate of Wear in Total Hip Replacement, *Clin. Orthoped.,* Vol 112, 1975, p 170
52. M.J. Griffith, M. Seidenstein, D. Williams, and J. Charnley, Socket Wear in Charnley Low Friction Arthroplasty of the Hip, *Clin. Orthoped.*, Vol 137, 1978, p 37-47
53. B.M. Wroblewski, 15 to 20 Year Results of the Charnley Low Friction Arthroplasty, *Clin. Orthopaed. Rel. Res.,* Vol 211, 1985, p 30-35
54. J. Livermore, D. Elstrup, and B. Morrey, Effect of Femoral Head Size on Wear of Polyethylene Acetabular Component, *J. Bone Joint Surg.,* Vol 72A, 1990, p 518-528
55. J.R. Atkinson, G.H. Isaac, D. Dowson, and B.M. Wroblewski, Laboratory Wear Tests and Clinical Observations of the Penetration of Femoral Heads Into Acetabular Cups in Total Replacement Hip Joints. III. The Measurement of Internal Volume Changes in Explanted Charnley Sockets After 2-16 Years In Vivo and the Determinations of Wear Factors, *Wear,* Vol 104, 1985, p 225-244
56. H.G. Willert, G.H. Buchhorn, and P. Eyerer, Ed., *Ultra-High Molecular Weight Polyethylene as Biomaterial in Orthopaedic Surgery,* Hogrefe & Huber, Göttingen, 1991, p 1-272
57. H. Oonishi, H. Aoki, and K. Sawai, Ed., *Bioceramics,* Vol 1, Ishiyaku Euro America, Tokyo, 1989, p 1-434
58. D. Dowson and N.C. Wallbridge, Laboratory Wear Tests and Clinical Observations of the Penetration of Femoral Heads Into Acetabular Cups in Total Replacement Hip Joints. I. Charnley Prostheses With

Polytetrafluoroethylene Acetabular Cups, *Wear,* Vol 104, 1985, p 203-215

59. J.P. Paul, Forces Transmitted by Joints in the Human Body, *Proc. Inst. Mech. Eng.,* Vol 181 (Part 31), 1967, p 8-15
60. N.C. Wallbridge and D. Dowson, The Walking Activity of Patients Fitted With Artificial Joints, *Eng. Med.,* Vol 11 (No. 2), 1982, p 95-96
61. D. Dowson, M.M. El-Hady Diab, B.J. Gillis, and J.R. Atkinson, Influence of Counterface Topography on the Wear of Ultra High Molecular Weight of Polyethylene Under Wet or Dry Conditions, *Polymer Wear and Its Control,* L.-H. Lee, Ed., Symp. Series 287, American Chemical Society, 1985, p 215-228
62. D. Dowson, S. Taheri, and N.C. Wallbridge, The Role of Counterface Imperfections in the Wear of Polyethylene, *Wear,* Vol 119, 1987, p 227-293
63. L. Zichner, U. Paschen, and M. Starker, In Vivo Wear of the Articulating Surfaces Made of Al_2O_3 Ceramics and Polyethylene in Total Hip End Prostheses, *Ultra-High Molecular Weight Polyethylene as Biomaterial in Orthopaedic Surgery,* H.G. Willert, G.H. Buchhorn, and P. Eyerer, Ed., Hogrefe & Huber, Göttingen, 1991, p 148-151
64. G. Langer, S. Blumentritt, and J. Babisch, The Wear Properties of Polyethylene Acetabular Cups With Respect to Various Femoral Components in Vivo, *Ultra-High Molecular Weight Polyethylene as Biomaterial in Orthopaedic Surgery,* H.G. Willert, G.H. Buchhorn, and P. Eyerer, Ed., Hogrefe & Huber, Göttingen, 1991, p 152-153
65. A. Egli, B.G. Weber, H. Sieber, M. Semlitsch, and E. Dörre, Experience With the Pairing of Polyethylene/Ceramic Materials in Hip Endoprostheses, *Ultra-High Molecular Weight Polyethylene as Biomaterial in Orthopaedic Surgery,* H.G. Willert, G.H. Buchhorn, and P. Eyerer, Ed., Hogrefe & Huber, Göttingen, 1991, p 154-158
66. H. Oonishi, H. Igaki, and Y. Takayama, Comparisons of Wear of UHMW Polyethylene Sliding Against Metal and Alumina in Total Hip Prostheses, *Bioceramics,* Vol 1, H. Oonishi, H. Aoki, and K. Sawai, Ed., Ishiyaku Euro America, Tokyo, 1989, p 272-277
67. T. Ohashi, S. Inoue, K. Kaikawa, K. Ibaragi, T. Tada, M. Oguchi, T. Arai, and K. Kondo, The Clinical Wear Rate of Acetabular Component Accompanied With Alumina Ceramic Head, *Bioceramics,* Vol 1, H. Oonishi, H. Aoki, and K. Sawai, Ishiyaku Euro America, Tokyo, 1989, p 278-283
68. K. Okumura, T. Yamamuro, T. Kumar, T. Nakamura, and M. Oka, Socket Wear in Total Hip Prosthesis With Alumina Ceramic Head, *Bioceramics,* Vol 1, H. Oonishi, H. Aoki, and K. Sawai, Ed., Ishiyaku Euro America, Tokyo, 1989, p 284-289
69. D. Dowson, Bio-Tribology of Natural and Replacement Synovial Joints, *Biomechanics of Diarthroidal Joints,* Vol II, C. Van Mow, A. Ratcliffe, and S.L.-Y. Woo, Ed., Springer-Verlag, 1990, p 305-345
70. D.C. Fricker, Friction When Femoral Prosthesis Heads Slide in Acetabular Cups, *Ceramics in Substitutive and Reconstructive Surgery,* P. Vincenzini, Ed., Elsevier, Amsterdam, 1991, p 207-215
71. A. Unsworth, D. Dowson, V. Wright, and D. Koshal, The Frictional Behaviour of Human Synovial Joints. Part II. Artificial Joints, *J. Lubr. Technol. (Trans. ASME),* Vol 97 (No. 3), 1974, p 377-382
72. J.O'Kelly, A. Unsworth, D. Dowson, B. Jobbins, and V. Wright, Pendulum and Simulator for Studies of Friction in Hip Joints, *Evaluation of Artificial Joints,* D. Dowson and V. Wright, Ed., Biological Engineering Society, 1977, p 19-20
73. A. Unsworth, M.J. Pearcy, E.F.T. White, and G. White, Frictional Properties of Artificial Hip Joints, *Eng. Med.,* Vol 17, 1988, p 101-104
74. J.R. Cooper, D. Dowson, and J. Fisher, Birefringent Studies of Polyethylene Wear Specimens and Acetabular Cup, *Wear,* Vol 151, 1991, p 391-402
75. D.F. Gibbons, Use of Polarized Light Microscopy for the Evaluation of UHMW-PE Components, *Ultra-High Molecular Weight Polyethylene as Biomaterial in Orthopaedic Surgery,* H.G. Willert, G.H. Buchhorn, and P. Eyerer, Ed., Hogrefe & Huber, Göttingen, 1990, p 60-62
76. J.A. Davidson and F.S. Georgette, "State of the Art Materials for Orthopaedic Prosthetic Devices," Technical Paper EM87-122, Society of Manufacturing Engineers, 1987, p 1-26
77. T.A. Andrew, J.P. Flanogan, M. Gerundini, and R. Bombelli, The Isoelastic, Non Cemented Total Hip

Arthroplasty--Preliminary Experience With 400 Cases, *Clin. Orthopaed. Rel. Res.,* No. 206, 1986, p 127-138

78. D. Dowson, Are Our Joint Replacement Materials Adequate?, *The Changing Role of Engineering in Orthopaedics*, Institution of Mechanical Engineers, London, 1989, p 1-5

Friction and Wear of Dental Materials

John M. Powers, University of Texas Health Science Center; Stephen C. Bayne, University of North Carolina

Introduction

HUMAN HARD DENTAL TISSUES (enamel, dentin, and cementum), restorative materials (metals, ceramics, and polymers), and dental instruments (burs, curettes, and endodontic files) used to abrade tooth structure are all subject to friction and wear processes. Table 1 summarizes some typical dental materials susceptible to wear; Table 2 presents typical mechanical properties of these same materials.

Table 1 Simplified composition or microstructure of some typical dental materials susceptible to wear and friction

Material	Simplified composition or microstructure
Alloys	
Alloy crown	Au-Ag-Cu (single phase), Au-Pd, Ni-Cr
Amalgam (admixed)	Ag_2Hg_3, Cu_6Sn_5, Ag_3Sn-Cu_3Sn-Zn (unreacted), Ag-Cu (unreacted)
Amalgam (unicompositional)	Ag_2Hg_3, Cu_6Sn_5, Ag_3Sn-Cu_3Sn-Zn (unreacted)
Endodontic instruments	Stainless steel, carbon steel
Orthodontic wire	Austenitic stainless steel, Ni-Ti-Co, Ti-Mo-Zr-Sn
Ceramics	
Cements:	
Zinc oxide-eugenol	Zinc eugenolate and zinc oxide (unreacted)
Zinc phosphate	Zinc phosphate and zinc oxide (unreacted)
Glass-ceramic inlays and crowns	Glass with crystalline alumina or fluoride-mica glass ceramic
Gypsum dies	$CaSO_4 \cdot 2H_2O$
Porcelain (PFM)	Pottassium-sodium alumino-silicate glasses
Porcelain denture teeth	Quartz particles in a glass matrix
Polymers	
Cements:	
Glass ionomer	Acrylic or maleic acid copolymer and calcium aluminosilicate matrix with aluminosilicate glass (unreacted) or silver-glass cermet (unreacted)
Zinc polyacrylate	Acrylic polymer and zinc oxide (unreacted)
Composites (hybrid)	Aromatic or urethane diacrylate matrix with silanated filler (quartz, borosilicate glass, lithium aluminum silicate, barium aluminum silicate, or colloidal silica)
Composites (microfilled)	Aromatic or urethane diacrylate matrix with colloidal silica
Denture acrylics	Poly(methyl methacrylate) with pre-polymerized PMMA or rubber-reinforced acrylic beads
Pit and fissure sealants	Aromatic or urethan diacrylate matrix with silanated filler
Plastic denture teeth	Poly(methyl methacrylate) matrix and beads

Source: Ref 1

Table 2 Typical properties of dental materials

Materials	Knoop hardness, kg/mm^2	Yield strength		Ultimate tensile strength		Ultimate compressive strength	
		MPa	**ksi**	**MPa**	**ksi**	**MPa**	**ksi**
Alloys							
Alloy crown (Au-Ag-Cu)	125 (HV)	220	32	425	62	. . .	. . .
Amalgam							
Admixed	. . .	. . .	. . .	40	5.8	340	49
Unicompositional	. . .	. . .	. . .	50	7.3	450	65
Orthodontic wires							
Stainless steel	540	1500	218	2000	290	. . .	. . .
Ni-Ti-Co	. . .	340	49	. . .	. . .	. . .	. . .
Ti-Mo-Zr-Sn	. . .	960	139	. . .	. . .	. . .	. . .
Ceramics							
Cements							
Zinc oxide-eugenol	. . .	. . .	. . .	5	0.7	40	5.8
Zinc phosphate	40	. . .	. . .	14	2.0	170	25
Human teeth							
Dentin	70	170	25	50	7.3	300	44
Enamel	340	350	51	10	1.5	380	55
Glass-ceramic inlays	360	. . .	. . .	. . .	. . .	. . .	. . .
Gypsum dies	65	. . .	. . .	8	1.2	80	11.6
Porcelain (PFM)	460	. . .	. . .	25	3.6	150	22
Porcelain denture teeth	460	. . .	. . .	. . .	. . .	. . .	. . .
Polymers							
Cements							
Glass ionomer	. . .	. . .	. . .	4	0.6	125	18
Zinc polyacrylate	. . .	. . .	. . .	10	1.5	70	10.2
Composites							
Hybrid	55	. . .	. . .	50	7.3	275	40
Microfilled	30	. . .	. . .	40	5.8	290	42
Denture acrylics	20	25	3.6	50	7.3	75	10.9
Pit and fissure sealant	25	. . .	. . .	30	4.4	170	25
Plastic denture teeth	20	55	8	. . .	. . .	76	11.0

Source: Ref 1

This article will review the friction and wear of various dental materials that have been studied by fundamental wear measurements, simulated service wear measurements, and clinical measurements (Ref 2, 3). Correlations of properties such as hardness, fracture toughness, and wear will be described and wear mechanism such as sliding adhesive wear, two-body abrasion, three-body abrasion, erosion, and fatigue identified.

Human Dental Tissues

The destruction of hard dental tissues by wear is categorized conveniently as physiologic wear, pathologic wear, prophylactic wear, and finishing procedure wear (Ref 4). In of these types, the actual wear situation may vary considerably because of different substrates, opposing wear surfaces, lubrication systems involved, and third-body abrasives. An overview of different circumstances is provided in Table 3.

Table 3 Classification of wear situations in dentistry

Intraoral wear event	Type of wear	Lubricant	Substrate	Opponent	Abrasive
Physiologic causes of wear					
Noncontact wear	3-body	Saliva/food	Tooth/restoration	. . .	Food
Direct contact wear	2-body	Saliva	Tooth/restoration	Tooth/restoration	. . .
Sliding contact wear	2-body	Saliva	Tooth/restoration	Tooth/restoration	. . .
Pathologic causes of wear					
Bruxism	2-body	Saliva	Tooth/restoration	Tooth/restoration	. . .
Xerostomia	2-body	. . .	Tooth/restoration	Tooth/restoration	. . .
Erosion	. . .	Saliva	Tooth/restoration	. . .	. . .

Unusual habits	2-body	Saliva	Tooth/restoration	Foreign body	. . .
Prophylactic causes of wear					
Toothbrush and dentifrice	3-body	Water	Tooth/restoration	Toothbrush	Dentifrice
Prophylactic pastes	3-body	Water	Tooth/restoration	Polishing cup	Pumice
Scaling and cleaning instruments	2-body	Saliva	Tooth/restoration	Instruments	. . .
Cutting, finishing, polishing					
Cutting burs/diamonds	2-body	Water	Tooth/restoration	Bur	. . .
Finishing burs	2-body	Water	Tooth/restoration	Bur	. . .
Polishing pastes	3-body	Water	Tooth/restoration	Polishing cup	Abrasive slurry

Physiologic Wear

Physiologic wear, or attrition (Ref 5), is caused by processes involving sliding contact wear, contact wear (impact without sliding), and noncontact wear from food abrasion alone. Sliding contact wear produces the most prominent effects during masticatory function. An example of severe attrition is shown in Fig. 1. Attrition occurs because of function (Ref 6) and occurs only where the opposing teeth come into contact (Ref 7), so it can be distinguished from erosion (see the discussion of "Pathologic Wear" given below).

Fig. 1 Severe occlusal wear resulting from dental attrition. Source: Ref 4

Attrition in primitive man was often severe (with pupal exposure) owing to the nature of the aboriginal diet of tough meat and sandy, fibrous plants, as studies on both skeletal and living representatives of aboriginal populations have attested (Ref 8, 9, 10).

Fundamental research on the degree and types of attrition found in the teeth of ancient, primitive, and modern populations has dealt mainly with the wear planes produced on molars surfaces (Ref 11, 12, 13) and with the effects of attrition on facial height (Ref 14, 15, 16, 17) and on the position of the temporomandibular joint (Ref 16, 18). It is apparent that not all people wear on their teeth the same way. The patterns produced are frequently characteristic of ethnic variations.

Improved tooth care and dietary habits have lessened considerably the occurrence of attrition in modern populations. For additional information on the effects of food, degree of function, and age on attrition, see Ref 19, 20, 21, 22, 23, 24, 25.

Pathologic Wear

This cause of wear can be particularly destructive to individual teeth or the entire dentition. Xerostomia and bruxism are the most frequently reported pathologic causes of wear.

Xerostomia. This condition results in dryness of the oral cavity and causes brittleness of the teeth. It has been observed in women during and after menopause (Ref 26). Investigation of causes of abnormal tooth wear must take salivary factors into consideration (Ref 27, 28). Mucin-based saliva substitutes lubricate with values comparable to whole human saliva, whereas substitutes based on carboxymethylcellulose do not appear to lubricate well (Ref 29).

Bruxism. This condition is a nonfunctional mandibular movement that is manifested by occasional or habitual grinding or clenching of the teeth (Ref 30). The major effects of severe bruxism can be tooth wear and accelerated alveolar bone loss. An example of occlusal wear resulting from bruxism is shown in Fig. 2. This abnormal wear rapidly removes the cusps of teeth. Wear takes place mainly on incisal edges of upper and lower anterior teeth. With time, edges become highly polished and flattened. In the posterior teeth, wear appears as small saucer-like excavations.

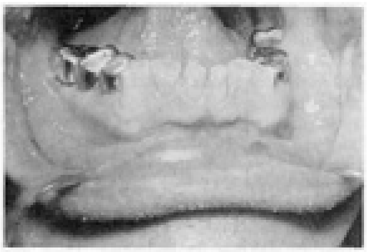


Fig. 2 Occlusal wear resulting from bruxism. Source: Ref 4

Erosion. This condition is the chemical weakening of human enamel resulting from acid decalcification. The pH affects the rate of decalcification for cleaned, and mechanically abraded enamel (Ref 31). Chemical decalcification may be caused by environmental pollutants contaminating air and/or saliva of patients. Accelerated wear is observed in people employed in occupations (for example, mining, or sulfuric acid production) where an unusual or severe atmospheric environment exists (Ref 32).

Unusual Habits. People who grasp needles or nails with their teeth or smoke pipes may also exhibit localized pathologic wear.

Prophylactic Wear

Toothbrush and Dentifrice. Oral hygiene is necessary for maintaining a healthy mouth and for social acceptance. The emphasis in the study of wear of the dentition by toothbrush and dentifrice has been the elimination of overly abrasive dentifrice components. Cervical abrasion resulting from improper and excessive toothbrush and dentifrice use is shown in Fig. 3.

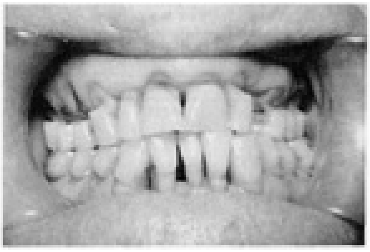


Fig. 3 Cervical abrasion resulting from excessive toothbrush and dentifrice use. Source: Ref 4

The primary function of a dentifrice is to clean and polish the surfaces of the teeth accessible with a toothbrush. During cleaning, extraneous debris and deposits need to be removed from the tooth surface. These deposits listed in order of increasing difficulty of removal are (Ref 33): food debris, plaque (a soft, mainly bacterial film), acquired pellicle (a proteinaceous film of salivary origin), and calculus.

The ideal abrasive should exhibit a maximum cleaning efficiency with minimum tooth abrasion. In addition, a dentifrice should polish the teeth. Highly polished teeth are not only aesthetically desirable, but they may also be less receptive to the retention of deposits (Ref 34).

Typical dentifrice abrasives include: calcium carbonate, dibasic calcium phosphate dihydrate, anhydrous dibasic calcium phosphate, tricalcium phosphate, calcium sulfate, calcium pyrophosphate, insoluble sodium metaphosphate, and hydrated alumina (Ref 35).

Selection of a dentifrice by a dentist for a patient should be based on: (1) degree of staining, (2) force exerted on the brush, (3) method of brushing, and (4) amount of exposed dentin and cementum. The Council on Dental Therapeutics of the American Dental Association published information on the abrasivity of dentifrices in 1970 (Ref 36).

Prophylactic Paste. A dental prophylactic paste should be sufficiently abrasive to remove effectively exogenous stains, pellicle, *materia alba,* and oral debris from the tooth surface without causing undue abrasion to the enamel, dentin, or cementum. Polymeric materials, such as denture base and artificial tooth resins, composite restorations, and pit and fissure sealants, are particularly susceptible to abrasion because of their low hardness. The undesirable results of wear can be a reduction in anatomic contours and increased surface roughness.

Abrasives in commercial prophylactic pastes include: recrystallized kaolinite, silicon dioxide, calcined magnesium silicate, diatomaceous silicon dioxide, pumice, sodium-potassium-aluminum silicate, and zirconium silicate (Ref 37).

Cutting, Finishing, and Polishing Wear

Tooth structure and restorative dental materials are routinely reshaped and smoothened using special instruments for cutting and finishing. A highly polished surface is then produced by treatment with polishing pastes containing alumina or diamond abrasive particles less than 1 μm in size.

Wear Studies

Traditional wear theory divides observed wear into categories of adhesive, abrasive, corrosive, and fatigue wear; however, predictions of these wear models depend on the materials behaving in a relatively brittle fashion. Most dental materials under intraoral circumstances do not behave in this way; therefore, it is difficult to rank dental materials performance. Most wear tests have not faithfully predicted clinical performance.

Wear information on dental materials has been collected from fundamental studies with simple laboratory tests, simulation studies with customized machines, and clinical studies. Unfortunately, the fundamental laboratory tests and the simulation studies have not had much success in correlating with observed clinical wear.

Fundamental Laboratory Studies

In a single-pass sliding technique, fluorapatite single crystals served as a simple model system for human enamel, which is composed of hydroxyapatite. The wear and friction of fluorapatite single crystals under conditions of single- and multiple-pass sliding with a diamond hemisphere (360 μm in diameter) can be evaluated by interpretation of tangential force, track width, and surface failure classification data (Ref 38, 39, 40). A failure classification scale (Fig. 4) includes:

- *Class 1:* entirely ductile
- *Class 2:* mostly ductile with some tensile cracking
- *Class 3:* essentially tensile cracking
- *Class 4:* mostly tensile cracking with chevrons (chipping)
- *Class 5:* chevrons

Examples of Classes 1, 3, and 5 are shown in Fig. 5.

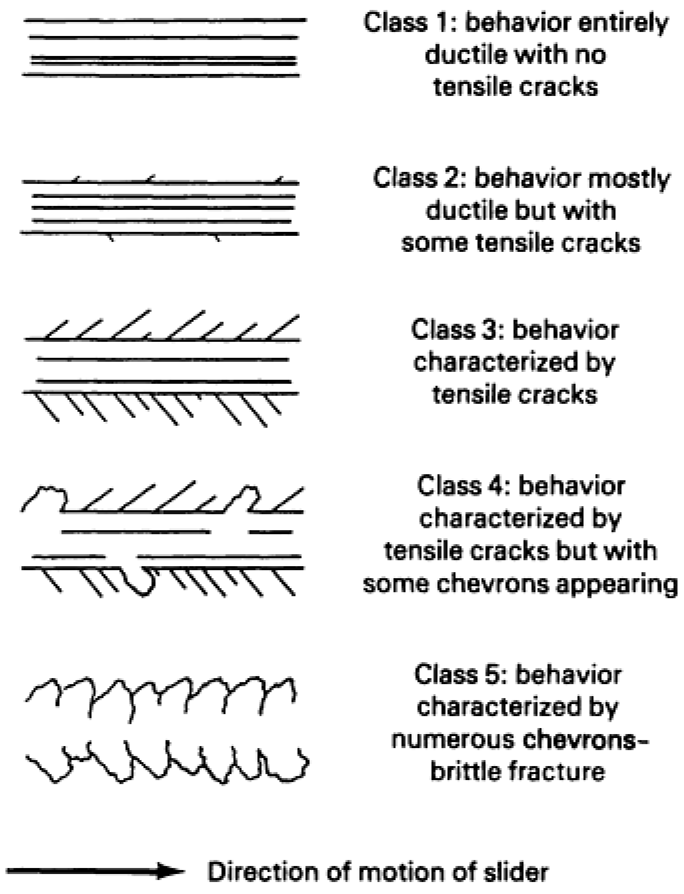


Fig. 4 Failure classification scale. Source: Ref 40

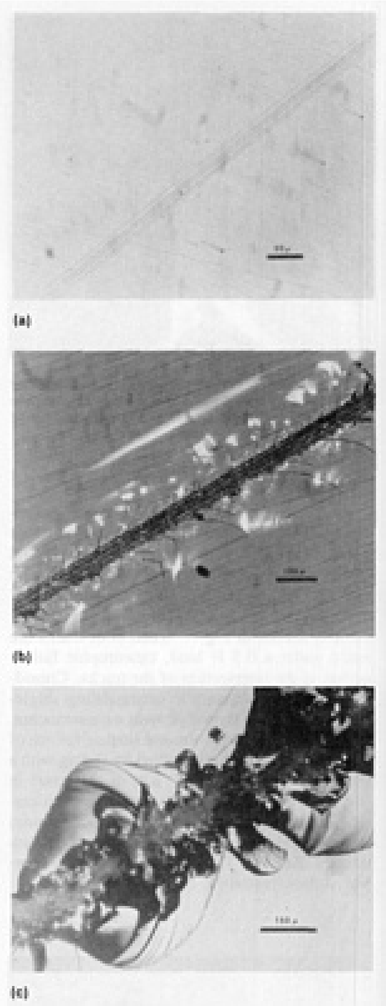


Fig. 5 Examples of surface failure of fluorapatite single crystals. (a) Class 1. (b) Class 3. (c) Class 5. Source: Ref 40

The failure of fluorapatite at a 0.1 N load for single-pass sliding in the <2110> direction is essentially ductile and progresses toward brittle failure as the load is increased (Ref 41). At 0.5 N loads and higher, failure is characterized by chevron formation. Track width follows an exponential function, whereas the tangential force (friction) increases linearly with normal load (Ref 41). The coefficient of friction is not a perfect indicator of wear. The track width data indicate that the principal mechanisms for the accommodation of strain are elastic deformation and cracking. Sliding in the <0110> direction results in slightly lower friction but substantially increased surface damage (Table 4).

Table 4 Influence of sliding direction and environment on friction and wear properties of fluorapatite single crystals for single-pass sliding on the basal plane

Condition	Coefficient of friction (β)	Ductile-to-brittle transition load, N
<2110> direction in air	0.22	0.75
<0110> direction in air	0.19	0.15
<2110> direction in water	0.24	0.18
<2110> direction in dimethylformamide	0.24	0.62

Source: Ref 41, 42

The ductile-to-brittle transition for sliding of diamond on fluorapatite occurs at higher normal loads in air and dimethylformamide than in water (Table 4) (Ref 42). The lowering of the transition in water is explained on the basis of

surface hardening as a result of the interaction of polar water molecules or their dissociation products and charged near-surface species. The interaction results in pinning of dislocations and a reduction in the ability of the lattice to accommodate strain by slip and thus effectively lowers the stress required to cause fracture.

Single-pass sliding by itself cannot completely describe the wear of fluorapatite single crystals (Ref 43). Wear tracks for a single and double pass on the same track in the opposite direction under a 0.7 N load are shown in Fig. 6. The effect of sliding a second pass across a wear track is shown in Fig. 7. When both tracks are made under a 0.5 N load, catastrophic failure occurs at the intersection of the tracks. Considerable care is necessary in extrapolating single-pass wear data to repetitive wear measurements.

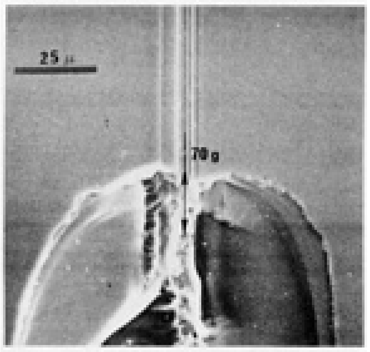


Fig. 6 Wear tracks on a fluorapatite single crystal for a single and double pass on the same track in the opposite direction under a 0.7 N load. Source: Ref 43

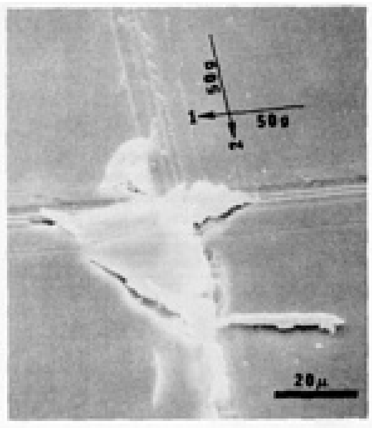


Fig. 7 Intersecting wear tracks on a fluorapatite single crystal under a 0.5 N load. Source: Ref 43

The frictional behavior and surface failure of human enamel has been studied by sliding with a diamond hemisphere (360 μm in diameter) in water (Ref 44, 45). Similar tests have been carried out on a sintered hydroxyapatite ceramic, which approximates the properties of human enamel (Ref 46). Table 5 compares the properties of these materials.

Table 5 Properties of a sintered hydroxyapatite ceramic compared with human enamel

Property	HAP-60K-1200C	Human enamel
Compressive strength, MPa (ksi)	380 (55)	400 (58)
Young's modulus, GPa (psi × 10^6)	120 (17.4)	80 (11.6)
Knoop hardness, kg/mm^2	450	340
Density, g/cm^3	3.1	2.96
Coefficient of friction	0.24	0.36

Linear coefficient of thermal expansion, 10^{-6}/ °C	9.2-11.8	11.4

Source: Ref 46

Simulation Studies

Chewing Machines. An early attempt to study attrition used a machine capable of simulating the actions of the human mandible during chewing to produce a variety of wear patterns (Ref 47). Much more sophisticated chewing machines have been developed to examine dental composite wear and are discussed later.

Brushing Machines (Dentifrices). Three main methods used to determine the loss of hard tooth tissue from brushing are: (1) measuring the amount of tooth tissue abraded from an irradiated tooth by the concentration of radioactive phosphorus in the wear debris (Ref 36, 48, 49, 50, 51), (2) determining the change in profile of samples (Ref 52), and (3) measuring the change in reflectance of the surface of tooth structure (Ref 53).

Dentifrices used in 1942 were compared to a calcium carbonate standard (Ref 54, 55). Abrasion was found to occur 25 times faster on dentin and 35 times faster on cementum than on the enamel tips of cusps. Abrasive power (percent abrasion of a calcium carbonate control) increases linearly with particle size (Ref 56, 57). On the other hand, the polishing ratio of zirconium silicate increases with decreasing particle size, and particle size distribution is important (Ref 53). Increasing the load on the brush also increases enamel and dentin abrasion scores (Ref 50, 58).

The radiotracer method can measure wear rates of dental tissues by as few as one or two strokes of a brush on dentin with commercial dentifrices. Use of this method has led to the observation that wear of dental tissues is proportional to penetration hardness if hard abrasives are used (Ref 49).

Brushing Machines (Toothbrush). A number of studies have attempted to determine the influence of the toothbrush and its variables on the wear of dental tissues. Plastic toothbrush bristles have little abrasive or polishing power (Ref 59, 60, 61, 62). Other studies have compared automatic versus hand toothbrushes (Ref 63, 64, 65). In general, the mechanical toothbrushes produce less abrasion of hard tooth tissue than simulated manual brushing, but the forces associated with manual brushing are usually larger.

Many variables of toothbrushes and dentifrices have been examined with brushing machines. (These include (Ref 4):

- *Dentifrice properties:* hardness, particle size, and particle distribution of the abrasive, and composition and concentration of remaining components
- *Toothbrush properties:* geometry, hardness, stiffness, and number of bristles
- *Substrate properties:* orientation, hardness, surface preparation
- *Testing conditions:* brush load, stroke length, stroke rate, number of strokes, and presence of saliva

Machines for Prophylaxis. Products containing quartz and pumice show higher cleansing values but generally result in greater abrasion to enamel and dentin (Ref 66). Abrasion of dentin by a pumice slurry is about 20 times greater than abrasion of enamel under standardized conditions (Ref 67). Increases in treatment time, load and cup speed cause linear increases in abrasion of irradiated human enamel and dentin (Ref 68). Commercial products containing calcined magnesium silicate and sodium-potassium-aluminum silicate show best polishing with low abrasivity (Ref 37). No clinical study has yet correlated the degree of abrasivity with any destructive effect on hard and soft tissues. American Dental Association Specification No. 37 includes a suggested abrasivity test that uses a radiotracer technique (Ref 69).

Cutting Machines. Cutting of human enamel by a high-speed diamond stone is enhanced by use of a solution of glycerol, ethanol, and water (2:1:2) when compared to cutting with water alone (Ref 70). Such a chemomechanical effect is also observed in the cutting of amalgam with diamonds and carbide burs and the cutting of composites with carbide burs. Cutting of hydroxyapatite blocks with autoclaved tungsten carbide burs is enhanced if the burs are dipped in sodium nitrate or commercial anticorrosive dips before autoclaving (Ref 71).

Clinical Studies

Clinical studies collect physiologic wear data produced by: (1) sliding contacts or direct contact alone (occlusal-contact-area wear) and (2) noncontact wear (contact-free wear), which is related to food abrasion. To effectively examine these

events, pathologic and prophylactic wear must be absent or controlled. In addition to the restorative treatment variables in clinical studies, there are also intraoral variables and patient factors that complicate clinical results and interpretations.

Occlusal-contact-area wear of human enamel has been measured with a computerized three-dimensional measuring technique on tooth replicas over a period of four years (Ref 72). The steady-state wear rates at enamel occlusal contact areas are about 29 μm/year for molars and about 15 μm/year for premolars. These data agree with earlier reports of 33 μm/year (Ref 73) and 41 μm/year (Ref 74).

Observations of orthodontic patients with arrested carious lesions indicate that functional wear and toothbrushing are responsible for the arrestment by disturbance and removal of bacterial deposits (Ref 75). Changes in surface enamel morphology after acid etching are also the result of abrasion rather than the precipitation of mineral from saliva (Ref 62).

Enamel undergoes physiologic wear but is routinely redeposited by nucleation and growth of new hydroxyapatite from the calcium phosphate present in saliva. This process helps to compensate for losses that occur during physiologic, prophylactic, and polishing wear. Dentifrice abrasivity of enamel as measured *in vivo* by a cellulose acetate replication technique is much lower than abrasion caused by pumice or zirconium silicate (Ref 76).

Dental Amalgam

Dental amalgam is an alloy that results when mercury is mixed with an alloy containing silver, tin, copper, and sometimes zinc (Table 1). Before it hardens, the freshly mixed mass of amalgam can be packed into a cavity prepared in a tooth. Amalgams are usually limited to replacement of tooth issue in the posterior teeth and often function in stress-bearing areas susceptible to occlusal wear. Some properties of dental amalgams are listed in Table 2.

Fundamental Laboratory Studies

Abrasion Tests. Two-body abrasion of dental amalgam has been measured using a Taber abrader (Ref 77), a silicon carbide two-body abrasion test (Ref 78), and a pin-on-disk test (Ref 79). With the Taber abrader, smearing of amalgam and clogging of the abrader wheel cause inconsistent ranking with clinical observations; clogging of the abrasive surface is avoided by abrading at a low load over a fresh abrasive surface on each pass. Ranking of amalgam and composite restorative materials with two-body abrasion tests is in better agreement with clinical observations than that done with three-body abrasion tests. Two-body abrasion test result on some typical dental amalgam alloys are given in Table 6. As indicated by these results, a dispersed high-copper amalgam (Dispersalloy) exhibited better resistance to silicon carbide two-body abrasion than spherical low-copper amalgams.

Table 6 Material loss on abrasion of dental amalgams

Material	Material loss, 10^{-4} mm^3/mm of travel(a)	
	24 h	**1 month**
Spher-a-Caps	7.0	7.0
New True Dentalloy	6.5	6.3
Dispersalloy	5.6	4.9

(a) Load: 0.17 MPa.

The pin-on-disk test (Ref 79) utilizes a cylindrical sample of enamel to rub on a rotating disk of amalgam. Measurements of wear rate are possible, but transfer of material from the disk to the pin confuses interpretation of the results.

Single-Pass Sliding. The wear of dental amalgam also has been studied by single- and double-pass sliding with a diamond hemisphere (360 μm in diameter) (Ref 80). The dispersed high-copper amalgam has the lowest values of tangential force and track width. The mode of surface failure under single- and double-pass sliding is ductile with no evidence of subsurface failure. Smearing of phases for both spherical and dispersed amalgams during sliding does occur, but the dispersed amalgam is more resistant to smearing. Cracks that occur at higher loads propagate around the stronger

phases. Wear is determined by resistance to penetration and by a ductile mode of surface failure over the load range tested.

Friction of dental amalgam is altered when any transfer of material from one member of the pair to the other member occurs (Ref 81). For example, when gold or dental composite slide against amalgam, amalgam material is transferred to the gold or composite surface and the friction then becomes that of amalgam on amalgam.

Fracture toughness, critical strain energy release rate, and critical stress intensity factor have been determined for several types of dental amalgams (Ref 82). Data are consistent with surface failure observed in single-pass wear studies.

An equation developed from single-pass studies (Ref 83) has been derived that relates the sliding frictional force (F) to normal load (N), fracture toughness, modulus of elasticity, yield strength, and slider diameter in the form of $F = KN^n$ (Table 7). The observed friction is caused primarily by plowing or deformation during single-pass sliding.

Table 7 Friction of dental materials as described by various properties including fracture toughness

Material	Modulus of elasticity		Yield strength		Fracture toughness, J/m^2	*n*	*K*
	GPa	**psi × 10^6**	**MPa**	**ksi**			
Amalgams							
Spherical, low-copper	12.9	1.87	142	20.6	117	1.17	0.211
Lathe-cut, low-copper	12.8	1.86	119	17.3	247	1.31	0.146
Admixed, high-copper	17.7	2.57	130	18.9	104	1.25	0.124
Composite	7.8	1.13	135	19.6	182	1.23	0.190
Unfilled resin							
Acrylic	2.0	0.29	104	15.1	382	1.54	0.141
Diacrylate	1.8	0.26	65	9.4	402	1.19	0.338

Source: Ref 83

Simulation Studies

Amalgam abrasion from dentifrices in brushing machines has been studied by profilometry and laser reflection techniques, which both indicate that amalgam wears about 10 times more rapidly than gold alloys under the same conditions of simulated brushing with normal dentifrices (Ref 84).

Clinical Studies

The physiologic wear rates of amalgam are 6 to 15 μm/year in contact-free areas and 28 to 58 μm/year in occlusal contact areas (Ref 85). This wear may be compensated by continual amalgam expansion that produces occlusal extrusion. Therefore, the rate of attrition of amalgam is usually not considered to be a clinical problem.

Amalgam degrades at the tooth-restoration interface by a process called marginal fracture, which is the result of electrochemical corrosion in the presence of direct-contact stresses (Ref 86). Direct-contact areas on amalgams that are not stabilized by contacts on enamel may also produce noticeable facets.

Composite Restorative Materials

Composite restorative materials consist of a cross-linked polymer matrix that is chemically bonded by coupling agents to the surfaces of dispersed silica-based filler particles (Table 1). Composite restorations have the appearance of natural tooth tissue and can be placed directly into a cavity preparation for *in situ* hardening. They are recommended for restorations where occlusal stress is minimal and appearance is crucial. Composites are also available for limited posterior use in areas of occlusal stress, but they are less durable than amalgam (Ref 87). Some properties are listed in Table 2.

Wear resistance of composite restorations is important for clinical longevity, esthetics, and resistance to dental plaque. The need for markedly improved wear resistance has been emphasized by recent literature reviews (Ref 88, 89, 90, 91). The reason that wear-resistant composites have not been developed is primarily because of the lack of understanding of the mechanisms of clinical wear. This problem is confounded by the lack of reliable and consistent clinical wear data.

Fundamental Laboratory Studies

The problem with all wear theories has been the lack of correlation of clinical wear results with laboratory properties. Mechanical properties do correlate with filler types and loading levels (Ref 92, 93, 94, 95, 96, 97), but they do not predict clinical performance. To bridge the gap, a wide range of tests have been performed. Some of these are described below.

Abrasion Tests. Taber abrasion (Ref 81), single-pass sliding on abrasive papers (Ref 98), pin and plate (Ref 99), pin and disk (Ref 100), metallographic polishing (Ref 101), toothbrushing (Ref 102), and oscillating wear (Ref 103) tests have shown very little or on success in explaining intraoral wear. With the Taber abrader, smearing of resin and clogging of the abrader wheel produced apparently better resistance for unfilled resins than for composites; these results are inconsistent in ranking with clinical observations.

A silicon carbide two-body abrasion test was developed to avoid clogging effects (Ref 78). This test provided some agreement with early clinical studies, but its primary usefulness was to examine the relative effects of different formulation variables.

Improvements in the durability of composites depend on a thorough understanding of the wear bahavior of the resin matrix (Ref 104). The resin with the lowest value of two-body abrasion also showed the lowest coefficient of friction and the most ductile behavior during single-pass sliding. (Table 8). Abrasion resistance was improved by finishing the cured surface before testing (Ref 105). Reduced wear of large-particle composites was related to the increased size (relative to the abrasive), increased hardness, and increased volume fraction of the filler particles (Ref 106, 107).

Table 8 Material loss on abrasion and single-pass wear characteristics of dimethacrylate resins

Material	Material loss, 10^{-4} mm^3/mm of travel[(a)]	Coefficient of friction (β)	Ductile-to-brittle transition load, N
Tetraethylene glycoldimeth-acrylate	22.0	0.45	7.0
Bisphenol A-bis ethylmeth-acrylate	17.7	0.61	5.0
Bisphenol A-bis (2-hydroxypropyl)methacrylate + ethylene glycol-dimethacrylate (1:1)	15.5	0.35	>10
Bisphenol A-bis ethylmeth-acrylate + octafluoro-1-pentyl-methacrylate (9:3)	19.1	0.61	4.0
Bisphenol A-bis ethylmeth-acrylate + octafluoro-1-pentyl-methacrylate (3:9)	32.2	1.48	3.0

Source: Ref 104

(a) Load: 0.18 MPa.

Pin-on-disk testers were popular at least in part become subsurface damage detected during microdefect analysis of a few composites recovered from *in vivo* service was apparently the same as the structure observed in composites subjected to pin-on-disk wear using stainless steel sliders (Ref 108).

Two-body pin-on-disk abrasion of composite veneering resins was shown to be much higher for rubbing contact against porcelain versus enamel, gold, or the same veneering resin (Ref 109). Combinations of composite with composite, silver-reinforced glass ionomer, or porcelain wore excessively (Ref 110).

The enthusiasm for abrasion testing produced a large amount of data, but actual clinical wear mechanisms often proved to be quite different.

Single-Pass Sliding. The wear of composites also has been studied by single- and double-pass sliding with a diamond hemisphere (360 μm in diameter) (Ref 111, 112). The surface failure for unfilled diacrylate resins was more severe than that observed for an unfilled acrylic resin. Addition of nonsilanated filler to the diacrylate resin increased resistance to penetration but did not dramatically change the mode of failure (Fig. 8). Diacrylate resins that contained silanated filler (commercial composites) were ductile in failure during sliding and showed the higher resistance to penetration. Damage was more severe for double-pass than for single-pass sliding. Single-pass data did not, however, provide an estimate of volume lost on repetitive wear of composites.

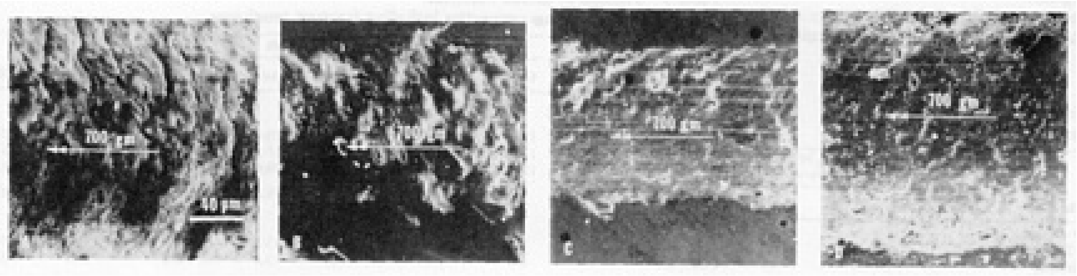


Fig. 8 Scanning electron micrographs of double-pass wear scars made under a normal load of 7 N of (A) unfilled diacrylate resin, (B) unfilled acrylic, (C) diacrylate resin with silanated filler, (D) diacrylate resin with nonsilanated filler. Source: Ref 112

Aging and Chemical Softening. The surface degradation of composites caused by accelerated aging in a weathering chamber was characterized by erosion of the resin matrices and exposure of filler particles (Ref 113). Differences in surface roughness and profile indicated that the aged composites were eroded at different rates. Surface crazing (Fig. 9) was observed for some aged composites. Surface degradation also resulted in color changes, particularly for large-particle composites but much less so for microfilled composites (Ref 114).

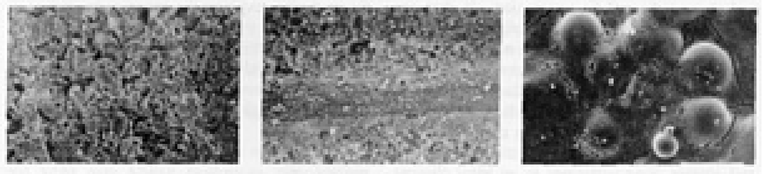


Fig. 9 Scanning electron micrographs of composites aged for 900 h showing surface crazing. Source: Ref 113

Single-pass sliding wear of aged, chemically cured, large-particle composites resulted in smaller track widths, lower tangential forces, easier dislodging of material, and more severe surface failure at lower normal loads than that of unaged composites (Ref 115). Ductile-to-brittle transitions occurred at lower normal loads (Table 9), and changes in morphology became more severe (Fig. 10) with increased aging (Ref 116).

Table 9 Influence of increased aging on single-pass wear transitions of composites under conditions of accelerated aging

Material	Ductile-to-brittle transition load, N		
	Unaged	**300 h**	**900 h**
Large-particle composite			
Concise	2.8	1.8	1.5
Microfilled composite			
Finesse	8.0	3.3	3.0
Isopast	4.3	4.3	4.3
Silar	7.0	6.0	2.0

Source: Ref 116

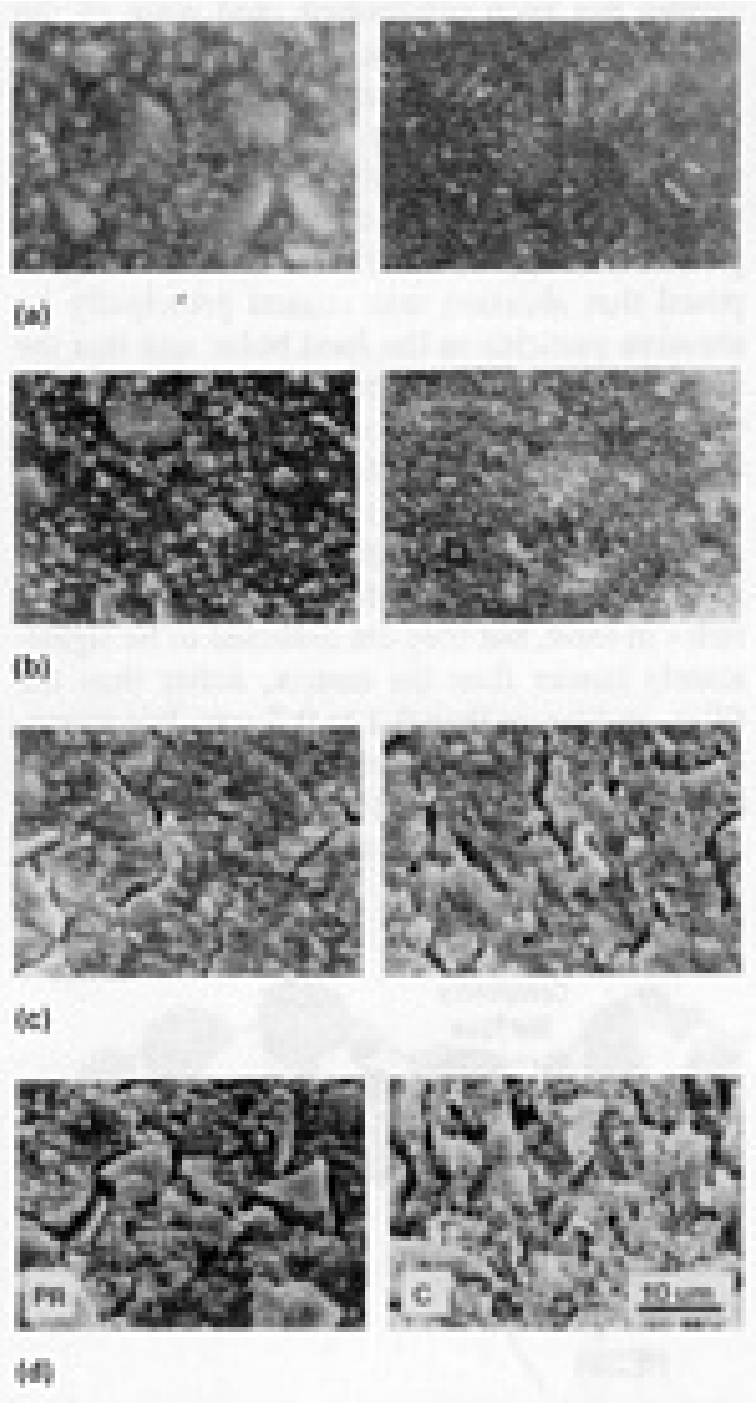


Fig. 10 Scanning electron micrographs of surface morphology of two large-particle composites at (a) 0, (b) 300, (c) 600, and (d) 900 h. Source: Ref 116

Preconditioning composites in food-simulating liquids (heptane and several ethanol/water solutions) has been shown to decrease the hardness with a corresponding increase in pin-on-disk wear (Ref 117). Swelling of the polymer matrix and surface damage occurs during preconditioning. Increasing the degree of cure of the matrix polymer may inhibit diffusion of penetrants, and additional cross-linking may reduce swelling and damage by solvents. Evidence of erosion has also been observed under *in vivo* conditions (Ref 118).

Fracture toughness, critical strain energy release rate (G_{Ic}), and critical stress intensity factor (K_{Ic}) have been determined for experimental and commercial composites (Ref 119). The commercial composite was less resistant to crack initiation and had a higher K_{Ic} than its unfilled diacrylate resin. Data were consistent with surface failure observed in single-pass wear studies of these resins.

Correlating Abrasion with Hardness and Tensile Data. Hardness and tensile strength of large-particle composites were not related to measured *in vitro* abrasion rates (Ref 120). However, clinical abrasion of microfilled composites and unfilled resins appeared related to indentation hardness (Ref 121).

Simulation Studies

Most simulation studies evaluate physiological wear. Few studies have dealt with prophylactic wear. Microfilled composites subjected to brushing abrasion with a dentifrice wore 5 to 10 times faster than large-particle composites (Ref 122). Abrasion slowed once the resin-rich layer of a composite was lost or when the composite was postcured by heat.

A limited study was reported in which composites were abraded by an artificial food bolus (Ref 123). Composites were placed along the edge of a rotating disk immersed in a millet seed suspension and ground against a stainless steel wheel. Although not perfect, the ranking of composites agreed with clinical studies.

Chewing machines have been developed to simulate the full range of intraoral wear events (Ref 124). To date, the results have only shown a limited agreement with clinical studies.

Clinical Studies

Many direct clinical evaluations of wear rely on the United States Public Health Service (USPHS) criteria for classifying the loss of anatomical form of posterior restorations (Ref 125). Unfortunately, the USPHS criteria alone are not sufficient to detect early loss of material in posterior composites (Ref 126, 127, 128, 129, 130, 131). Because of the crudeness of the USPHS scale, there was considerable impetus in the early 1980s to develop more precise indirect measurement techniques.

The first indirect method to become popular was designated the Leinfelder technique. It measured the loss of material using a stone cast poured from a clinical impression of the restoration to compare it to a series of calibrated stone casts (Ref 126, 130, 132). Other similar indirect cast methods have tired to refine the Leinfelder approach (Ref 133, 134, 135). Unfortunately, because of operational differences, these scales have been shown to underestimate actual wear by 50%.

The major advantage of indirect methods employing stone casts and evaluators is their inexpensive nature. All other methods, particularly computer digitization, are remarkably costly and thus impractical for even small clinical studies.

Electroplated or epoxy resin replicas of restorations have been prepared from impressions and studied by optical, scanning electron, or reflex microscopy; by Moiré techniques; or with computer digitization (Ref 86, 136). Wear has also been measured by monitoring the recession of resin on the surfaces of prominent filler particles by scanning electron microscopy of sequences of impressions of a restoration (Ref 137).

A less-complicated method of clinically evaluating restorations is to place them into denture teeth in a removable denture (Ref 138, 139, 140, 141, 142). The dentures permit direct measurements of the samples by profilometry or three-dimensional computerized surface digitization.

Because of the large number of uncontrollable clinical variables involved in these studies, there is a broad distribution of results at each recall. The results are different for molars and premolars. Differences in operators and techniques among clinical studies are often so great that the results are not comparable at all. The reported results are often misrepresented as wear rates rather than total wear. Wear does not occur in a linear fashion.

Published wear rates for composites are 12 to 79 μm/year in contact-free areas and 39 to 135 μm/year in occlusal contact areas (Ref 85). For autocured (Ref 143) and ultraviolet light-polymerized (Ref 144, 145) composites, wear has been shown to occur at decreasing rates during the first three years (Fig. 11). This same pattern of decreasing wear has now been reported for several newer visible light cured composites (Ref 146, 147, 148). For one composite, about 74% of the total wear over a three-year period occurred during the first year, followed by 19% in the second year, and 6% in the third year. This pattern is less obvious if the first six months of data are excluded. The wear rate for the next few years may appear almost constant (Ref 149). The rate reported for a few microfilled composites appeared to be nearly linear (Ref 85, 150), but early wear may have been hidden by beveling of the restoration margins. Recently, wear rates for posterior composites have been reported to drop to very low values after 5 to 8 years (Ref 145, 147, 148). Because of this overall decreasing rate of wear, it is hazardous to project long-term wear rates from any short period of time.

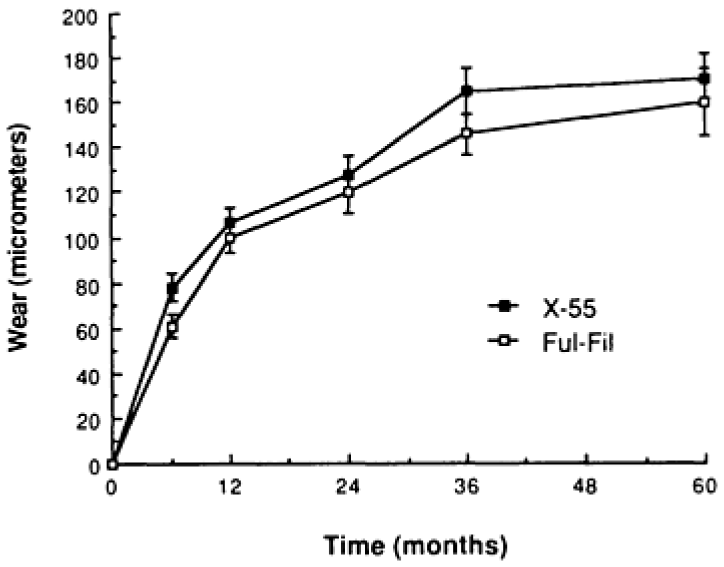


Fig. 11 Decreasing clinical wear rate with time. Source: Ref 148

Wear of posterior conventional composites involves exfoliation of the filler particles as the resin matrix is continually worn away (Ref 147, 151). This process gives the appearance of a restoration submerging below the surface (Fig. 12). The wear is reduced in the hybrid composites by reducing the mean filler particle size from a range of 30 to 50 μm to a range of 3 to 5 μm and by increasing the amount of filler from 75 to 86 wt% (Ref 128). The use of softer barium glass (HK 400) to replace quartz filler (HK 600) in the particle size range of 10 to 20 μm also results in better wear resistance, although the glass particles themselves show evidence of wear.

Fig. 12 Scanning electron micrograph of a composite restoration (Concise) with severe occlusal wear after a seven-year period. 10×. Source: Ref 188

Other factors leading to breakdown include the degradation of the silane coupling agent, which can cause microcracking of the resin (Ref 152) and the large mismatch in coefficient of thermal expansion between filler and resin (Ref 86, 118). Maximizing filler particle-to-particle contacts does not improve wear resistance (Ref 153). Resistance to wear is higher for unfinished composites than for those finished conventionally with a finishing bur and white stone, suggesting that finishing at least initially mechanically weakens the surface (Ref 154).

Another problem for clinical research is that different evaluation systems record different wear events. For the most part, the indirect method relies on relief occurring at margins to create contrasts that can be compared to standards. This method does not measure wear elsewhere on the occlusal surface. It is not clear where on the restoration direct clinical

evaluation measures wear, but it seems most likely to be at the marginal areas. Profiling methods measure the general occlusal surface very well but are least accurate at the margins (Ref 155, 156).

A further complication of clinical research studies is that there has always been the wide distribution of wear values for individual restorations at any recall period. After three years, the wear ratings for composites vary from 0 to 350 μm (Ref 148).

In an attempt to understand this variation, the factors contributing to intraoral wear have been collected into categories for statistical analysis (Ref 157). The categories include factors involving cavity preparation, restoration, manipulation, intraoral location, and the patient. Although not all of these factors are yet well known, it has been possible to identify the major ones.

Intraoral location is the most important factor. First molars wear more than second molars, second premolars, and first premolars, in that order. Restoration width is also important. There are minor effects from arch, gender, and complexity of the restoration. Differences in formulations of materials have secondary effects (Ref 158). Understanding these effects, it is possible to normalize differences among clinical studies to pool the clinical data for posterior composites and correct these variations to an ideal clinical population using Weibull analysis (Ref 159).

Numerous theories of wear have been proposed for dental composites based on their microstructures and resistance to microstructure. The analysis typically is presented in terms of the filler, coupling agent, or resin matrix being the weakest link (Fig. 13). The percent conversion of monomer to polymer during curing (Ref 160, 161, 162, 163, 164, 165, 166, 167), the depth of curing (Ref 168, 169, 170, 171, 172, 173, 174), the glass transition temperature of the polymer (Ref 175, 176), and the strength of the polymer (Ref 177, 178, 179, 180, 181, 182) have all been proposed as reasons that the matrix might be most susceptible to microfracturing and thus wear. Problems with coupling agent effectiveness and hydrolysis have been proposed as possible weak links (Ref 152). It has been hypothesized that the surfaces of composites absorb chemicals from food and become predisposed to microfracture, thus permitting wear (Ref 108, 117). Different wear mechanisms have been theorized for different composite types (Ref 149). None of these mechanisms has been established, and none of the explanations is predictive.

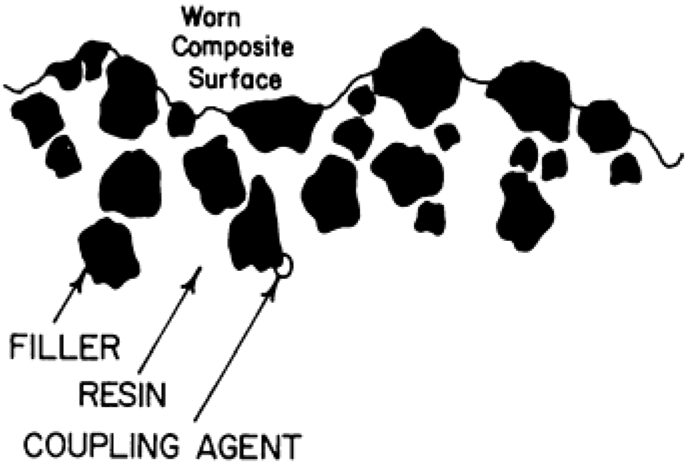


Fig. 13 Schematic of worn composite restoration indicating possible weak links

A key observation was made by examination of local patterns of wear on composite surfaces (Ref 121, 138). Wear of resin seemed to occur only when separation of filler particles was greater than approximately 0.1 μm. It was proposed that abrasion was caused principally by abrasive particles in the food bolus and that the smallest effective particles were larger than 0.1 to 0.2 μm. Therefore, composites with interparticle spacings less than that dimension were microscopically protected against wear. No information has been available about the size, amount, or abrasivity of the wear-producing particles in food, but they are assumed to be significantly harder than the matrix, softer than the filler, and larger than 0.1 to 0.2 μm. For microfills and many hybrids containing ≥20 vol% of 0.4 μm microfiller particles, the computed average filler particle separation is <0.2 μm. This theory readily explains the apparent resistance to wear of

those composites *in vivo.* Theoretical calculations of microfiller levels required for microprotection indicate that only a few percent are required. However, assuming that microfiller tends to be agglomerated, levels of 45 to 50% are necessary.

An extension of this argument has been that macroscopic protection would also exist for small cavity preparations. Increased exposure of the preparation walls shelters the adjacent composite and causes a decrease in the clinical wear rate. Most of the observed wear patterns and decreasing wear rates appear to be explainable on the basis of these two protection mechanisms alone (Fig. 14).

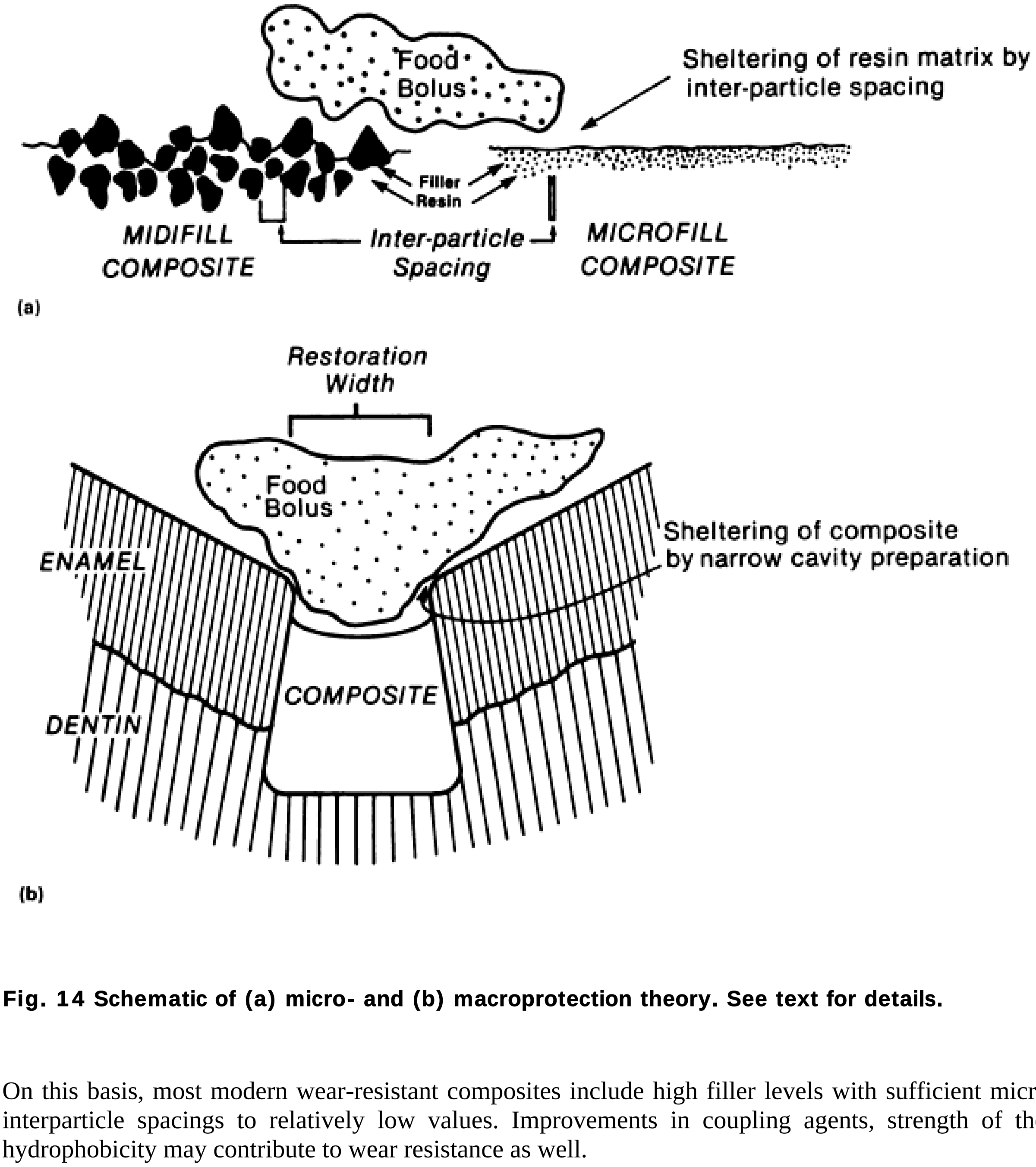


Fig. 14 Schematic of (a) micro- and (b) macroprotection theory. See text for details.

On this basis, most modern wear-resistant composites include high filler levels with sufficient microfiller to reduce the interparticle spacings to relatively low values. Improvements in coupling agents, strength of the resin matrix, and hydrophobicity may contribute to wear resistance as well.

Pit and Fissure Sealants

Pit and fissure sealant consist mainly of a polymer matrix with minor amounts of dispersed filler particles, primarily for coloration (Table 1). Sealants are designed to be more fluid than composites so they will penetrate the pits, fissures, and etched areas of enamel to produce macroscopic and microscopic mechanical retention. The purpose of sealants is to penetrate all cracks, pits, and fissures on the occlusal surfaces of deciduous or permanent teeth at risk to dental caries, to seal these areas, and to provide effective protection against caries-producing bacteria. Some properties of sealants are listed in Table 2.

Fundamental Laboratory Studies

Two-body abrasion of sealants has been measured using a silicon carbide two-body abrasion test (Ref 183). Weight-loss data range from 22×10^{-4} to 24×10^{-4} mm^3/mm of travel, which is characteristic of wear of diacrylate resins with little or no filler. The addition of 40 wt% quartz to a sealer did not affect its resistance to two-body abrasion.

The wear of sealants also has been studied by single-pass sliding with a diamond hemisphere (360 μm in diameter) (Ref 183). The sealant with 40 wt% quartz was more resistant to penetration and showed less surface damage from single-pass sliding than two unfilled diacrylate sealants.

Clinical Studies

Direct clinical observations of 205 restorations after 4 years has shown that physiologic wear of sealants in the non-stress bearing areas of pits and fissures is only slight (7%) (Ref 184). Wear on other occlusal surface areas where sealant may be applied inadvertently is apparently irrelevant, because these areas are naturally self-cleansing by abrasion from food. Therefore, the wear of sealants in that regard is clinically unimportant.

Dental Cements

Various modified zinc oxide-eugenol and zinc polyacrylate dental cements are used for temporary fillings (Table 1). Glass ionomers and silver-reinforced glass ionomers may be used for certain permanent resolutions not subject to high occlusal forces. Representative mechanical properties of cements are listed in Table 2.

Table 10 Material loss on abrasion of dental cements used as temporary filling materials

Material	Material loss, 10^{-4} mm^3/mm of travel[(a)]
Zinc oxide-eugenol cements	
Tem Pak	29.1
B and T	22.9
IRM	18.6
EBAC	15.8
Zinc polyacrylate cements	
PCA Cement	5.0
Durelon	4.9

Source: Ref 185

(a) Load: 0.036 MPa.

Fundamental Laboratory Studies

Two-body abrasion of temporary filling materials has been measured using a silicon carbide two-body abrasion test (Ref 185). The materials are ranked in an order that agrees with clinical observations. Two zinc polyacrylate cements (base consistency) had rates of abrasion much lower than modified zinc oxide-eugenol cements (Table 10).

The two-body abrasion of early glass ionomer cements used for Class V restorations is similar to that of modified composites (Table 11) (Ref 186). Pin-on-disk wear of silver-reinforced glass ionomer restorative materials is less than that for conventional glass ionomers (Ref 187). Incorporation of silver appears to provide lubrication; however, the incidence of catastrophic failure during sliding is only reduced slightly.

Table 11 Material loss on abrasion of Class V restorative materials

Material	Material loss, 10^{-4} mm^3/mm of travel[(a)]
Glass ionomer cements	
ASPA	22.9
Fuji	17.0
Composites	
Cervident	14.7
Enamelite 500	19.5

Source: Ref 186

(a) Load: 0.18 MPa.

Clinical Studies

Clinical investigations of glass ionomer restorations show increasing surface roughness and volume loss (Ref 188). In the absence of occlusal forces, glass ionomers demonstrate some erosion, but a good internal bond is still maintained between the glass particles and the gel matrix (Ref 118).

Noble and Base Metal Alloys

Numerous alloys based on gold-silver-copper, gold-palladium, or nickel-chromium are commercially available to make crowns to restore part or all of the coronal portion of a tooth. These alloys are used primarily in the posterior portion of the mouth where high strength is required and cosmetic appearance is secondary. They are presumed to be resistant to wear and have always been assumed to produce little or no wear of opponent teeth. Thus, very little research has been conducted. Crowns of porcelain fused to metal are often used to combine the aesthetics of porcelain with the strength and fit of a cast crown. In these situations, the abrasion characteristics of the metal alloy are supplanted by the properties of the porcelain.

Clinical Studies

From the limited data available, it appears that wear rates of gold alloys and porcelain are the same when opposing porcelain-fused-to-metal crowns, as shown in three patients with known bruxing and wear problems (Ref 189).

Porcelain and Plastic Denture Teeth

Porcelain or plastic denture teeth are fabricated at factories rather than in the dental office or dental laboratory. Their representative composition is reported in Table 1. Controversy continues among dentists as to the greater benefits of porcelain versus plastic teeth for prosthetic appliances (complete dentures, partial dentures). Shortly after plastic teeth were introduced (*circa* 1940), research reports condemning the choice of acrylic teeth for dentures also started appearing (Ref 190, 191). Since that time, acrylic teeth have been substantially improved and are now the prosthetic tooth material of choice for many dentists.

Fundamental Laboratory Studies

Two-body abrasion of acrylic teeth against sandblasted glass or with a Taber abrader has demonstrated that acrylic wears much faster than 22 carat gold, porcelain, and enamel, but only about 5% more rapidly than dentin (Ref 81, 192, 193). Under intermittent sliding, the rougher the porcelain, the more rapid was the wear of opposing gold and enamel (Ref 194).

The frictional behavior and surface failure of acrylic and porcelain denture teeth have been studied by single- and double-pass sliding with a diamond hemisphere (360 μm in diameter) in water and saliva (Ref 195, 196). Coefficients of friction are higher for acrylic teeth than for porcelain teeth (Table 12). Acrylic teeth show ductile behavior under sliding with ductile-to-brittle transitions occurring about 5 N. Porcelain teeth show brittle behavior under sliding with ductile-to-brittle transitions occurring between 1.5 and 2.5 N. Differences in wear from sliding in water or saliva are not significant. Acrylic teeth show greater deviations from a Hertzian model.

Table 12 Friction and wear properties of acrylic and porcelain denture teeth for single-pass sliding

Material	Coefficient of friction (β)	Ductile-to-brittle transition load, N
Acrylic teeth		
Dentsply		
Dentin	0.62	. . .
Enamel	0.62	4.8
Myerson		
Dentin	0.55	. . .
Enamel	0.54	5.6
Porcelain teeth		
Dentsply		
Dentin	0.24	. . .
Enamel	0.24	2.6
Myerson		
Dentin	0.26	. . .
Enamel	0.26	2.4

Source: Ref 195, 196

The friction of various prosthetic tooth materials varies from wet to dry conditions (Ref 81, 197). When the materials in the two-body pair both involve polar bonding, the presence of water or other polar liquids increases the friction. However, when one of the solids involves nonpolar bonding, water either has no effect or acts as a mild lubricant. Lower loads on acrylic surfaces produce lower friction as a result of the ductile behavior of acrylic. The friction coefficient of acrylic/porcelain combinations (β = 0.30) is less than acrylic/acrylic (β = 0.37) and porcelain/porcelain (β = 0.51) combinations.

Simulation Studies

Chewing machines have traditionally been employed to evaluate abrasion resistance of prosthetic teeth. An articulator attached to a grinding machine was used as an early abrasion test (Ref 198). Dentures with full sets of acrylic teeth, porcelain teeth, or gold restorations were abraded against a denture with porcelain teeth in water, with and without an abrasive. Wear was measured as changes in vertical dimension of occlusion. Acrylic teeth showed heavy wear. Porcelain teeth withstood abrasion but were friable. Teeth with gold restorations wore severely when an abrasive was present. Similar tests with an instrument using a hemispherical pin on an elliptical disk (similar to a stomatognathic system) showed that acrylic/acrylic and porcelain/acrylic combinations had less wear than porcelain/porcelain combinations (Ref 199).

The results of different pairs of wearing surfaces have not been consistent for all wear machines. Results from several chewing machines showed that acrylic/acrylic combinations produced less wear than acrylic/porcelain (Ref 200) and that wear decreased as contacting surface area of the acrylic teeth increased (Ref 201). In other wear machine tests, less wear was observed for porcelain/acrylic combinations than with acrylic/acrylic or porcelain/porcelain combinations (Ref 202, 203).

Recent occlusal wear studies using an artificial mouth and computerized profilometry to measure wear have shown that denture teeth made from a highly cross-linked copolymer with an interpenetrating polymer network (IPN) were more wear resistant than traditional acrylic resin teeth (Ref 204). The IPN denture teeth do not appear to produce measurable wear in contact with human teeth.

Clinical Studies

Acrylic teeth, when studied as a means of obtaining balanced functional occlusion in partial dentures, showed no tendency to wear over a period of 6 to 24 months (Ref 205). Abrasion of acrylic teeth has been observed clinically in dentures of 9 of 12 patients (Ref 206). A five-year clinical appraisal of 114 denture patients with acrylic teeth showed a loss of occlusion in centric relation as a result of wear of acrylic posterior teeth (Ref 207). Porcelain teeth and variables

such as bone loss or varying occlusal concepts were not evaluated. The results of a survey of 77 dentists were that porcelain/acrylic combinations appear to wear less than acrylic on acrylic (Ref 208).

Techniques involving removable sections of a fixed prosthesis along with replicas have been used to study wear in a male patient with extreme bruxing habits (Ref 209). Light-cured composite, gold alloy, porcelain, and heat-cured acrylic restorations generally wore more when opposed by porcelain or acrylic denture teeth (Table 13). Wear mechanisms observed included: combined tribochemical wear and fatigue for heat-cured acrylic, fatigue wear for light-cured composite and porcelain, and abrasion and fatigue wear for gold alloys. Wear of both porcelain and cross-linked resin teeth was mainly via fatigue. Abrasive wear occurred in the presence of hard particles (Ref 210). Microfilled composite resin teeth appeared to wear by fatigue and tribochemical wear (Ref 189).

Table 13 Loss of materials (mm^3/month) caused by acrylic or porcelain denture teeth on opposing restoration

Restorations	Opposing denture teeth	
	Acrylic	**Porcelain**
Heat-cured acrylic	3.9	8.3
Light-cured composite	1.9	5.4
Porcelain	1.0	1.1
Type III/IV gold alloy	0.6	1.2

The average vertical loss of denture teeth appears to be 0.1 mm or less per year for acrylic resin teeth (Ref 211, 212) and less for porcelain teeth (Ref 211). Clinical wear of a single acrylic tooth for one year has been measured from computer-generated data obtained from a reflex microscope and is 7.2 mm^3 (Ref 213).

Denture Acrylics

Dentures are made primarily from poly(methyl methacrylate), vinyl-acrylic, or rubber-reinforced acrylic polymers (Table 1). Their main use is to support plastic or porcelain artificial teeth. Some typical properties are listed in Table 2.

Fundamental Laboratory Studies

Two-body abrasion of heat-cured and self-cured acrylic denture resins correlates with scratch measurements and values of flexural modulus of elasticity (Ref 214). The abrasive wear rate of denture resins was much higher than that of composites, alloys, or teeth (Ref 215).

Dental Feldspathic Porcelain and Ceramics

Dental porcelain is primarily a glass with either some dispersed leucite or alumina crystals, or fluoride-mica crystals (Table 1). Porcelain has excellent aesthetic properties and resists wear extremely well. Some representative properties are reported in Table 2.

Fundamental Laboratory Studies

The frictional behavior and surface failure of dental feldspathic porcelain has been studied by single-pass sliding tests using a diamond hemisphere (360 μm in diameter) in air and water (Ref 216). The friction of as-glazed porcelain is higher and the surface damage is more extensive in water than in air (Table 14). Crack initiation within porcelain occurs at flaws such as scratches or residual pores, and these are propagated by elastic stresses associated with normal intraoral loads. The higher friction in bulk water is attributed to increased cracking and reduced amounts of recoverable elastic strain energy. Coating the porcelain with gold or chromium makes the friction independent of environment. Gold coating reduces the friction, but chromium coating does not. Chromium coating reduces the extent of damage from sliding compared to the as-glazed porcelain.

Table 14 Friction and wear properties of dental feldspathic porcelain for single-pass sliding

Condition	Coefficient of friction (β)	Ductile-to-brittle transition load, N
As-glazed		
Air	0.18	5.9
Water	0.23	4.9
Gold coated		
Air	0.14	5.9
Water	0.14	4.9
Chromium coated		
Air	0.22	5.9
Water	0.22	5.9

Source: Ref 216

Simulation Studies

Simulated occlusal wear studies in an artificial mouth using enamel occluding on porcelain has demonstrated a high coefficient of wear for dental porcelain and suggests that an abrasive wear mechanism is involved (Ref 217).

Die Materials (Stone, Resin, and Metal)

The most common die material is calcium sulfate dihydrate (gypsum), but dies may be fabricated from epoxy resin or electroplated with silver or copper (Table 1). A dental die is a replica of the hard or soft tissues, or both, and must be strong and resistant to abrasion because it is subjected to the stresses of carving and finishing a restoration. Some representative properties are listed in Table 2.

Fundamental Studies

Two-body abrasion of stone, resin, and metal-plated dies is shown in Table 15. The metal-plated dies have the highest hardness and lowest loss in two-body abrasion. Although the dental stone is harder than the resins, the surface morphology of the stone (Fig. 15) makes it less abrasion resistant.

Table 15 Material loss on abrasion and surface hardness of dental die materials

Material	Material loss, 10^{-4} mm^3/ mm of travel	Knoop surface hardness, kg/mm^2
Dental stone		
Duroc	14.8	54
Duroc with hardener	16.3	62
Resin		
Die-Met	2.8	25
Epoxydent	4.8	25
Metal		
Copper-plated	1.5	134
Silver-plated	0.9	128

Source: Ref 218

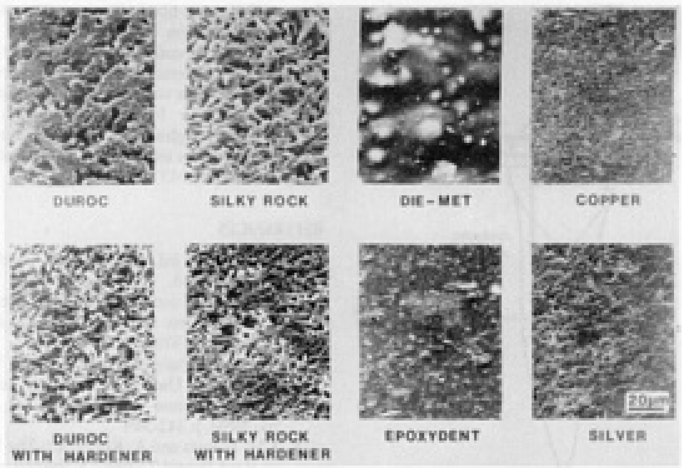


Fig. 15 Scanning electron micrographs of stone, resin, and metal-plated dies. Source: Ref 218

The surface failure of stone, resin, and metal-plated dies has been studied by single-pass sliding with a diamond hemisphere (360 μm in diameter) (Ref 218). The metal-plated dies are ductile up to a normal load of 10 N, whereas the resin and stone dies shown brittle failure above 1 to 2 N.

Endodontic Instruments

Endodontic instruments such as files and reamers are manufactured from stainless steel or carbon steel wires by various combinations of machining and twisting (Table 1). These instruments are used to clean and shape the root canal before it is sealed.

Simulation Studies

The cutting mechanism of a reamer on dentin involves compression and chipping (Ref 219). Triangular reamers cut deeper than rectangular ones. The smaller the angle of the cutting edge, the greater the cutting ability (Ref 220).

Cutting in bovine bone as measured by a drill press apparatus is more efficient with triangular than square reamers (Ref 221). Instruments appear to lose sharpness because of deformation or fracture of the blade.

Cutting in bovine bone or acrylic with a push-pull stroke in a linear motion also can be measured for various instruments (Ref 222, 223). Chips of acrylic appear to impair cutting efficiency. Reduced cutting efficiency has been observed with bovine bone.

Cutting ability of K-type stainless steel files was measured by depths of cut in poly(methyl methacrylate) on an instrument that controls the force on the files and the length and number of pull strokes (Ref 224). Variations in cutting are observed around the circumference of individual files and among files of the same size and brand (Ref 224, 225). Dry heat and salt sterilization has no effect on cutting ability of stainless steel files, but autoclave sterilization causes a reduction in their cutting ability (Ref 226). Sodium hypochlorite, hydrogen peroxide, and ethylene-diaminetetraacetic (EDTA) acid-urea peroxide irrigants cause a decrease in cutting ability, but a saline irrigant has no effect. Wear of K-type files under these conditions was not observed.

Cross-sectional design and flute design are more important variables in cutting than rake angle, wear resistance, capabilities for chip removal, and mode of use (Ref 227). Cutting efficiency of the tip of a file is better than the flute portion (Ref 228). In constricted canals (0.33 mm, or 0.013 in., in diameter), tip angles of files of 60 to 90° are most effective; whereas in larger canals (0.40 mm, or 0.016 in., in diameter), tip angles of files of 40 to 49° are most effective. The tip geometry, however, exerts a greater influence over cutting efficiency than tip length or angle. The pyramidal design is most efficient, whereas the conical tip is least efficient.

Periodontal Instruments

High-quality cutting edges of periodontal instruments are essential for effective subgingival scaling and root planning in periodontal therapy. Cuttability is affected primarily by cutting forces, chip thickness ratios, chip form, and tool wear (Ref 229, 230, 231). Edge quality may be evaluated by the angle between the two edge-forming contiguous surfaces, the smoothness of the edge, edge sharpness or dullness, and the presence or absence of metallic projections (Ref 232).

A sharp curette is the most proficient means for removing calculus and for achieving an acceptable, smooth root surface in periodontal therapy (Ref 233). A dull instrument leaves behind smeared deposits on the root surface, and a damaged curette causes heavy scratches.

Fundamental Laboratory Studies

Laboratory cutting tests of whale bone indicate that high-speed steel results in more-balanced chip formation than cemented carbide, better cutting forces than stainless steel and cemented carbide, less tool wear than stainless steel, and more ideal chip form than cemented carbide (Ref 231).

Simulation Studies

Root planning of extracted single-rooted teeth mounted in mannikin jaws indicates that stainless steel curettes show edge deformation after only 15 strokes (Ref 234). High-carbon steel curettes are more resistant to wear than stainless steel curettes (Ref 232).

Orthodontic Wires

Space closure and canine retraction in continuous arch wire techniques require sufficient force to overcome frictional forces between the bracket attached to the tooth and the arch wire (Fig. 16). Excessive wire/bracket friction may result in loss of anchorage or binding accompanied by little tooth movement. Factors that may influence friction are engagement of the arch wire in brackets that are out of alignment, ligatures pressing the arch wire against the base of the slot, active torque in rectangular wires, and bodily tooth movement in which the tipping tendency is resisted by two-point contact between the bracket and arch wire (Ref 235).

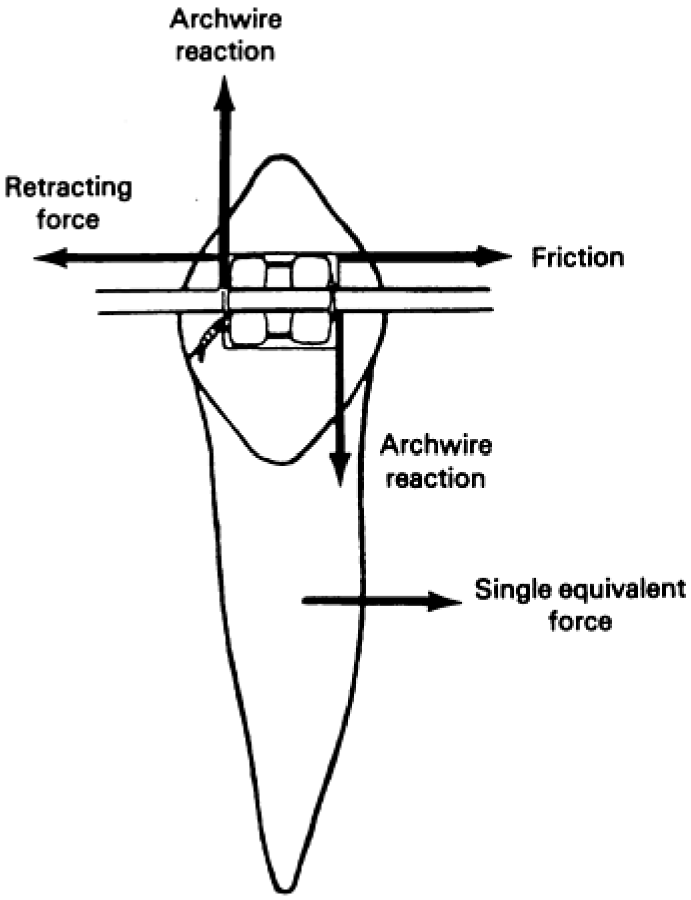


Fig. 16 Forces acting on a tooth/bracket system during translation. Source: Ref 235

Simulation Studies

Arch wire/bracket friction is affected by wire size, angulation, ligation force, bracket size, and wire type (Ref 236, 237, 238). Kinetic coefficients of friction of stainless steel, beta-titanium, nickel-titanium, and cobalt-chromium arch wires sliding on stainless steel or Teflon polymer increase with normal load (Table 16) and increase in the presence of artificial saliva as compared to dry conditions (Ref 239). Stainless steel and beta-titanium wires show wear tracks after sliding under low normal loads (40 to 80 N).

Table 16 Coefficient of friction of orthodontic wire/material combinations for a normal load of 40 and 80 N under wet conditions

Material	Coefficient of friction, 40 N	Coefficient of friction, 80 N
Stainless steel/stainless steel	0.035	0.085
Beta-titanium/stainless steel	0.025	0.122
Cobalt-chromium/stainless steel	0.041	0.122
Nickel-titanium/stainless steel	0.044	0.115
Stainless steel/Teflon	0.003	0.013

Source: Ref 239

There appears to be agreement that nickel-titanium wires have higher wire/bracket friction than stainless steel wires in zero torque/zero angulated brackets (Ref 240). At bracket angulations >3°, nickel-titanium wires have lower wire/bracket friction than stainless steel (Ref 238, 241). Under conditions of two-point contact, friction is inversely proportional to bracket width and is less for stainless steel wires than for nickel-titanium or beta-titanium wires (Ref 235). Surface roughness of nickel-titanium or beta-titanium wires may explain their relatively high friction (Ref 242).

References

1. R.G. Craig, Ed., *Restorative Dental Materials,* 8th ed., C.V. Mosby, 1989
2. R.G. Craig and J.M. Powers, Wear of Dental Tissues and Materials, *Int. Dent. J.,* Vol 26, 1976, p 121-133
3. M.Z.A.M Sulong and R.A. Aziz, Wear of Materials Used in Dentistry: A Review of the Literature, *J. Prosthet. Dent.,* Vol 63, 1990, p 342-349
4. J.M. Powers and A. Koran, III, The Wear of Hard Dental Tissue--A Review of the Literature, *J. Mich. Dent. Assoc.,* Vol 55, 1973, p 119-126
5. R.W. Leigh, *Dental Pathology of Aboriginal California,* University of California Press, 1928, p 408
6. M. Klatsky, Dental Attrition, *J. Am. Dent. Assoc.,* Vol 26, 1939, 73-84
7. H.L. Tollens, Atypical Tooth Erosion in a Pygmy Skull, *Dent. Abstr.,* Vol 5, 1960, p 668
8. P.O. Pederson, Investigations into Dental Conditions of about 3000 Ancient and Modern Greenlanders, *Dent. Rec.,* Vol 58, 1938, p 191-198
9. C.E. Snow, Indian Knoll Skeletons of Site Oh 2, Ohio County, Kentucky, *U. Kent. Reports in Anthrop.,* Vol 4, 1948, p 382-554
10. J.E. Anderson, Human Skeletons of Tehuacan, *Science,* Vol 148, 1965, p 496-497
11. L.H.D. Buxton, The Teeth and Jaws of Savage Man, *Trans. Brit. Soc. Study of Orthod.,* Vol 1916-1920, 1920, p 79-88
12. F.St.J. Steadman, Malocclusion in the Tasmanian Aborigines, *Dent. Rec.,* Vol 57, 1937, p 213-249
13. M.S. Goldstein, Dentition of Indian Crania from Texas, *Am. J. Phys. Anthrop.,* Vol 6, 1948, p 63-84
14. A.E. Webster, The Effect of Time and Wear on the Human Teeth, *Dent. Rec.,* Vol 38, 1919, p 331-335
15. G.G. Philippas, Effects of Function on Healthy Teeth: Evidence of Ancient Athenian Remains, *J. Am. Dent. Assoc.,* Vol 45, 1952, p 443-453
16. L. Lysell and R. Filipsson, A Profile Roentgenologic Study of a Series of Medieval Skulls from Northern Sweden, *Dent. Abstr.,* Vol 3, 1958, p 663
17. T. Murphy, Compensatory Mechanisms in Facial Height Adjustment to Function Tooth Attrition, *Austral. Dent. J.,* Vol 4, 1959, p 312-323
18. T. Murphy, Mandibular Adjustment to Functional Tooth Attrition, *Austral. Dent. J.,* Vol 3, 1958, p 171-178
19. D.K. Whittaker, T. Molleson, A.T. Daniel, J.T. Williams, P. Rose, and R. Resteghini, Quantitative Assessment of Tooth Wear, Alveolar-Crest Height and Continuing Eruption in a Romano-British Population, *Archs. Oral Biol.,* Vol 30, 1985, p 493-501
20. A.G. Comuzzie and D.G. Steele, Enlarged Occlusal Surfaces on First Molars Due to Severe Attrition and Hypercementosis: Examples from Prehistoric Coastal Populations of Texas, *Am. J. Phys. Anthropol.,* Vol 78, 1989, p 9-15
21. G.G. Philippas, Influence of Occlusal Wear and Age on Formation of Dentin and Size of Pulp Chamber, *J. Dent. Res.,* Vol 40, 1961, p 1186-1198
22. G.G. Philippas and E. Applebaum, Age Factor in Secondary Dentin Formation, *J. Dent. Res.* Vol 45, 1966, p 778-789
23. G.G. Philippas and E. Applebaum, Age Changes in the Permanent Upper Lateral Incisor, *J. Dent. Res.,* Vol 46, 1967, p 1002-1009
24. G.G. Philippas and E. Applebaum, Age Change in the Permanent Upper Canine Teeth, *J. Dent. Res.,* Vol 47, 1968, p 411-417
25. M.F. Teaford, A Review of Dental Microwear and Diet in Modern Mammals, *Scanning Microsc.,* Vol 2, 1988, p 1149-1166
26. W. Muller, Xerostomia, *Deut. Zahnarztebl.,* Vol 12, 1958, p 378-380
27. E. Zaus and G.W. Teuscher, Report on Three Cases of Congenital Dysfunction of the Major Salivary Glands, *J. Dent. Res.,* Vol 19, 1940, p 326

28. G.E. Carlsson, A. Hugoson, and G. Persson, Dental Abrasion and Alveolar Bone Loss in the White Rat. I. Effect of Ligation of the Major Salivary Gland Ducts, *Odont. Rev.,* Vol 16, 1965, p 308-316
29. M.N. Hatton, M.J. Levine, J.E. Margarone, and A. Aguirre, Lubrication and Viscosity Features of Human Saliva and Commercially Available Saliva Substitutes, *J. Oral. Maxillofac. Surg.,* Vol 45, 1987, p 496-499
30. S.C. Nadler, The Effects of Bruxism, *J. Periodont.,* Vol 37, 1966, p 311-319
31. J. Steel and R.C. Browne, Effect of Abrasion upon the Acid Decalcification of the Teeth, *Brit. Dent. J.,* Vol 94, 1953, p 285-288
32. L. Enbom, T. Magnusson, and G. Wall, Occlusal Wear in Miners, *Swed. Dent. J.,* Vol 10, 1986, p 165-170
33. C. Dawes, G.N. Jenkins, and C.H. Tonge, The Nomenclature of the Integuments of the Enamel Surface of Teeth, *Brit. Dent. J.,* Vol 115, 1963, p 65-68
34. R.W. Phillips and G. Van Huysen, Dentifrices and the Tooth Surface, *Am. Perf.,* Vol 50, 1948, p 33-41
35. S.D. Gershon, H.H. Pokras, and T.H. Rider, Dentifrices, *Cosmetics, Science and Technology,* E. Saragin, Ed., New York Interscience, 1957, p 296-353
36. American Dental Association, Council on Dental Therapeutics, Abrasivity of Current Dentifrices, *J. Am. Dent. Assoc.,* Vol 81, 1970, p 1177-1178
37. M.S. Putt, C.J. Kleber, and J.C. Muhler, Enamel Polish and Abrasion by Prophylaxis Pastes, *Dent. Hygiene,* Vol 56 (No. 9), 1982, p 38-43
38. J.M. Powers and R.G. Craig, Wear of Fluorapatite Single Crystals: I. Method for Quantitative Evaluation of Wear, *J. Dent. Res.,* Vol 51, 1972, p 168-176
39. J.M. Powers and R.G. Craig, Wear of Fluorapatite Single Crystals: II. Frictional Behavior, *J. Dent. Res.,* Vol 51, 1972, p 605-610
40. J.M. Powers and R.G. Craig, Wear of Fluorapatite Single Crystals: III. Classification of Surface Behavior, *J. Dent. Res.,* Vol 51, 1972, p 611-618
41. J.M. Powers, K.C. Ludema, and R.G. Craig, Wear of Fluorapatite Single Crystals: IV. Influence of Sliding Direction on Frictional Behavior and Surface Failure, *J. Dent. Res.,* Vol 52, 1973, p 1019-1025
42. J.M. Powers, K.C. Ludema, and R.G. Craig, Wear of Fluorapatite Single Crystals: V. Influence of Environment on Frictional Behavior and Surface Failure, *J. Dent. Res.,* Vol 52, 1973, p 1026-1031
43. J.M. Powers, K.C. Ludema, and R.G. Craig, Wear of Fluorapatite Single Crystals: VI. Influence of Multiple-Pass Sliding on Surface Failure, *J. Dent. Res.,* Vol 52, 1973, p 1032-1040
44. J.M. Powers, R.G. Craig, and K.C. Lubdema, Frictional Behavior and Surface Failure of Human Enamel, *J. Dent. Res.,* Vol 52, 1973, p 1327-1331
45. K.H.R. Wright, The Abrasive Wear Resistance of Human Dental Tissues, *Wear,* Vol 14, 1969, p 263-284
46. H.M. Rootare, J.M. Powers, and R.G. Craig, Sintered Hydroxyapatite Ceramic for Wear Studies, *J. Dent. Res.,* Vol 57, 1978, p 777-783
47. C.L. Brace and S. Molnar, Experimental Studies in Human Tooth Wear: I. *Am. J. Phys. Anthrop.,* Vol 27, 1967, p 213-222
48. R.J. Grabenstetter, R.W. Broge, F.L. Jackson, and A.W. Radike, The Measurement of the Abrasion of Human Teeth by Dentifrice Abrasion: A Test Utilizing Radioactive Teeth, *J. Dent. Res.,* Vol 37, 1958, p 1060-1068
49. K.H.R. Wright and J.I. Stevenson, The Measurement and Interpretation of Dentifrice Abrasiveness, *J. Soc. Cos. Chem.,* Vol 18, 1967, p 387-407
50. G.K. Stookey and J.C. Muhler, Laboratory Studies Concerning the Enamel and Dentin Abrasion Properties of Common Dentifrice Polishing Agents, *J. Dent. Res.,* Vol 47, 1968, p 524-532
51. J.J. Hefferren, A Laboratory Method for Assessment of Dentifrice Abrasivity, *J. Dent. Res.,* Vol 55, 1976, p 563-573
52. H. Ashmore, N.J. Van Abbe, and S.J. Wilson, The Measurement *in vitro* of Dentin Abrasion by Toothpaste, *Brit. Dent. J.,* Vol 133, 1972, p 60-66
53. G.K. Stookey, J.R. Hudson, and J.C. Muhler, Studies Concerning the Polishing Properties of Zirconium Silicate on Enamel, *J. Periodont.,* Vol 37, 1966, p 200-207

54. M.L. Tainter and S. Epstein, A Standard Procedure for Determining Abrasion by Dentifrices, *J. Am. Col. Dent.,* Vol 9, 1942, p 353-379
55. S. Epstein and M.L. Tainter, Abrasion of Teeth by Commercial Dentifrices, *J. Am. Dent. Assoc.,* Vol 30, 1943, p 1036-1045
56. S. Epstein and M.L. Tainter, The Relationship of Particle Size and Other Properties of Dentifrice Ingredients to Toothbrush Abrasion of Enamel, *J. Dent. Res.,* Vol 22 1943, p 335-344
57. W.H. Bull, R.M. Callender, B.R. Pugh, and G.D. Wood, The Abrasion and Cleaning Properties of Dentifrices, *Brit. Dent. J.,* Vol 125, 1968, p 331-337
58. R.S. Manly and D.H. Foster, Importance of Factorial Design in Testing Abrasion by Dentifrices, *J. Dent. Res.,* Vol 46, 1967, p 442-445
59. R.S. Manly, Factors Influencing Tests on the Abrasion of Dentin by Brushing with Dentifrices, *J. Dent. Res.,* Vol 23, 1944, p 59-72
60. R.W. Phillips and M.L. Swartz, Effects of Diameter of Nylon Brushes on Enamel Surface, *J. Am. Dent. Assoc.,* Vol 47, 1953, p 20-26
61. R.S. Manly and F. Brudevold, Relative Abrasiveness of Natural and Synthetic Toothbrush Bristles on Cementum and Dentin, *J. Am. Dent. Assoc.,* Vol 55, 1957, p 779-780
62. F. Mannerberg, Appearance of Tooth Surface as Observed in Shadowed Replicas in Various Age Groups, in Long Term Studies, after Toothbrushing, in Cases of Erosion, and after Exposure to Citrus Fruit Juices, *Odont. Rev. (Malmo),* Vol 11, Supplement 6, 1960, p 1-116
63. J.H. Harrington and I.A. Terry, Automatic and Hand Toothbrushing Studies, *J. Am. Dent. Assoc.,* Vol 68, 1964, p 343-350
64. R.S. Manly, J. Wiren, P.J. Manly, and R.C. Keene, A Method for Measurement of Abrasion by Toothbrush and Dentifrice, *J. Dent. Res.,* Vol 44, 1965, p 533-540
65. D. McConnell and C.W. Conroy, Comparisons of Abrasion Produced by a Simulated Manual Versus a Mechanical Toothbrush, *J. Dent. Res.,* Vol 46, 1967, p 1022-1027
66. V.E. Whitehurst, G.K. Stookey, and J.C. Muhler, Studies Concerning the Cleaning, Polishing, and Therapeutic Properties of Commercial Prophylaxis Pastes, *J. Oral Ther. Pharm.,* Vol 4, 1968, p 181-191
67. G.K. Stookey, *In vitro* Estimates of Enamel and Dentin Abrasion Associated with a Prophylaxis, *J. Dent. Res.,* Vol 57, 1978, p 36
68. G.K. Stookey and B.R. Schemehorn, A Method for Assessing the Relative Abrasion of Prophylaxis Materials, *J. Dent. Res.,* Vol 58, 1979, p 588-592
69. Specification 37-1986, American National Standards Institute/American Dental Association, Approved 15 May 1986
70. J.A. von Fraunhofer, C.D. Givens, and T.J. Overmyer, Lubricating Coolants for High-Speed Dental Handpieces, *J. Am. Dent. Assoc.,* Vol 119, 1989, p 291-295
71. G.K. Johnson, F.U. Perry, and G.B. Pelleu, Jr., Effect of Four Anticorrosive Dips on the Cutting Efficiency of Dental Carbide Burs, *J. Am. Dent. Assoc.,* Vol 114, 1987, p 648-650
72. P. Lambrechts, M. Braem, M. Vuylsteke-Wauters, and G. Vanherle, Quantitative *in vivo* Wear of Human Enamel, *J. Dent. Res.,* Vol 68, 1989, p 1752-1754
73. F. Roulet, P. Mettler, and U. Friedrich, Ein klinischer Vergleich dreier Komposits mit Amalgam für Klasse-II-Füllungen unter besonderer Berücksichtigung der Abrasion, Resultate nach 2 Jähren, *Schweiz. Monatsschr. Zahnkeilkd.,* Vol 90, 1980, p 18-30
74. S. Molnar, J.K. McKee, I.M. Molnar, and T.R. Przybeck, Tooth Wear Rates among Contemporary Australian Aborigines, *J. Dent. Res.,* Vol 62, 1983, p 562-565
75. L. Holmen, A. Thylstrup, and J. Artun, Surface Changes during the Arrest of Active Enamel Carious Lesions *in vivo.* A Scanning Electron Microscope Study, *Acta Odontol. Scand.,* Vol 45, 1987, p 383-390
76. S.V. Brasch, J. Lazarou, N.J. Van Abbe, and J.O. Forrest, The Assessment of Dentifrice Abrasivity *in vivo, Brit. Dent. J.,* Vol 127, 1969, p 119-124
77. A.A. Lugassy and E.H. Greener, An Abrasion Resistance Study of Some Dental Resins, *J. Dent. Res.,* Vol

51, 1972, p 967-972

78. J.M. Powers, L.J. Allen, and R.G. Craig, Two-Body Abrasion of Commercial and Experimental Restorative and Coating Resins and an Amalgam, *J. Am. Dent. Assoc.,* Vol 89, 1974, p 1118-1122
79. J.M. Powell, R.W. Phillips, and R.D. Norman, *In vitro* Wear Response of Composite Resin, Amalgam and Enamel, *J. Dent. Res.,* Vol 54, 1975, p 1183-1195
80. J.C. Roberts, J.M. Powers, and R.G. Craig, Wear of Dental Amalgam, *J. Biomed. Mater. Res.,* Vol 11, 1977, p 513-523
81. E.W. Tillitson, R.G. Craig, and F.A. Peyton, Friction and Wear of Restorative Dental Materials, *J. Dent. Res.,* Vol 50, 1971, p 149-154
82. J.C. Roberts, J.M. Powers, and R.G. Craig, Fracture Toughness and Critical Strain Energy Release Rate of Dental Amalgam, *J. Mater. Sci.,* Vol 13, 1978, p 965-971
83. J.C. Roberts, J.M. Powers, and R.G. Craig, An Empirical Equation Including Fracture Toughness and Describing Friction of Dental Restorative Materials, *Wear,* Vol 47, 1978, p 138-146
84. G. Johannsen, G. Redmalm, and H. Ryden, Surface Changes on Dental Materials, *Swed. Dent. J.,* Vol 13, 1989, p 267-276
85. J.C. Mitchem and D.G.Gronas, The Continued *in vivo* Evaluation of the Wear of Restorative Resins, *J. Am. Dent. Assoc.,* Vol III, 1985, p 961-964
86. P. Lambrechts, G. Vanherle, M. Vuylsteke, and C.L. Davidson, Quantitative Evaluation of the Wear Resistance of Posterior Dental Restorations: A New Three-Dimensional Measuring Technique, *J. Dent. Res.,* Vol 12, 1984, p 252-267
87. American Dental Association, Council on Dental Materials, Instruments, and Equipment, Posterior Composite Resins, *J. Am. Dent. Assoc.,* Vol 112, 1986, p 707-709
88. D.F. Taylor, *Posterior Composites,* D.F. Taylor, 1984, p 360
89. R.A. Draughn, R.L. Bowen, and J.P. Moffa, Composite Restorative Materials, *Restorative Dental Materials,* Vol 1, J.A. Reese and T.M. Valega, Ed., Biddles Ltd., 1985, p 75-107
90. *International State-of-the-Art Conference on Restorative Dental Materials,* Conference Proceedings, National Institute of Health/National Institute of Dental Research, 1986
91. J.F. Roulet, *Degradation of Dental Polymers,* Karger, 1987
92. M. Braem, P. Lambrechts, V. Van Doren, and G. Vanherle, The Impact of Composite Structure on Its Elastic Response, *J. Dent. Res.,* Vol 65, 1986, p 648-653
93. R.A. Draughn, Influence of Filler Parameters on Mechanical Properties of Composite Restorative Materials, *J. Dent. Res.,* Vol 62 (Special Issue), Abstract 187, 1983, p 670
94. K.J. Soderholm, Relationship between Compressive Yield Strength and Filler Fractions in PMMA Composites, *Acta Odontol. Scand.,* Vol 40, 1982, p 145-150
95. J.H. Hembree, W. Fingar, and R.A. Draughn, *South Carolina Dent. J.,* Vol 33, 1975, p 43-47
96. O. Zidan, E. Asmussen, and K.D. Jorgensen, Tensile Strength of Restorative Resins, *Scand. J. Dent. Res.,* Vol 88, 1950, p 285-289
97. H. Oysaed and I.E. Ruyter, Composites for Use in Posterior Teeth: Mechanical Properties Tested under Dry and Wet Conditions, *J. Biomed. Mater. Res.,* Vol 20, 1986, p 261-271
98. A.C McLundie, C.J.W. Patterson, and D.R. Stirrups, Comparison of the Abrasive Wear *in vitro* of a Number of Visible Light-Cured Composite Resins, *Brit. Dent. J.,* Vol 159, 1985, p 182-185
99. A. Harrison and T.T. Lewis, The Development of an Abrasion Testing Machine for Dental Materials, *J. Biomed. Mater. Res.,* Vol 9, 1975, p 341-353
100. S.L. Rice, W.F. Bailey, S.F. Wayne, and J.A. Burns, Comparative *in vitro* Sliding-Wear Study of Conventional Microfilled and Light-Cured Composite Restoratives, *J. Dent. Res.,* Vol 63, 1984, p 1173-1175
101. K.D. Jorgensen, *In vitro* Wear Tests of Macro-Filled Composite Restorative Materials, *Aust. Dent. J.,* Vol 27, 1982, p 153-158
102. Y. Li, M.L. Swartz, R.W. Phillips, B.K. Moore, and T.A. Roberts, Effect of Filler Content and Size on

Properties of Composites, *J. Dent. Res.,* Vol 64, 1985, p 1396-1401

103. R.M. Pilliar, D.C. Smith, and B. Maric, Oscillatory Wear Tests of Dental Composites, *J. Dent. Res.,* Vol 63, 1984, p 1166-1172
104. J.M. Powers, W.H. Douglas, and R.G. Craig, Wear of Dimethacrylate Resins Used in Dental Composites, *Wear,* Vol 54, 1979, p 79-86
105. G.S. Wilson, E.H. Davies, and J.A. von Fraunhofer, Abrasive Wear Characteristics of Anterior Restorative Materials, *Brit. Dent. J.,* Vol 151, 1981, p 335-338
106. R.A. Draughn and A. Harrison, Relationship between Abrasive Wear and Microstructure of Composite Resins, *J. Prosthet. Dent.,* Vol 40, 1978, p 220-224
107. G.S. Wilson, E.H. Davies, and J.A. von Fraunhofer, Micro Hardness Characteristics of Anterior Restorative Materials, *Brit. Dent. J.,* Vol 148, 1980, p 37-40
108. J.E. McKinney and W. Wu, Relationship between Subsurface Damage and Wear of Dental Restorative Composites, *J. Dent. Res.,* Vol 61, 1982, p 1083-1088
109. L.I. Gallegos and J.I. Nicholls, *In vitro* Two-Body Wear of Three Veneering Resins, *J. Prosthet. Dent.,* Vol 60, 1988, p 172-178
110. A. Embong, J. Glyn Jones, and A. Harrison, The Wear Effects of Selected Composites on Restorative Materials and Enamel, *Dent. Mater.,* Vol 3, 1987, p 236-240
111. J.M. Powers, J.C. Roberts, and R.G. Craig, Surface Failure of Commercial and Experimental Restorative Resins, *J. Dent. Res.,* Vol 55, 1976, p 432-436
112. J.M. Powers, J.C. Roberts, and R.G. Craig, Wear of Filled and Unfilled Dental Restorative Resins, *Wear,* Vol 39, 1976, p 117-122
113. J.M. Powers and P.L. Fan, Erosion of Composite Resins, *J. Dent. Res.,* Vol 59, 1980, p 815-819
114. J.M. Powers, P.L. Fan, and C.N. Raptis, Color Stability of New Composite Restorative Materials under Accelerated Aging, *J. Dent. Res.,* Vol 59, 1980, p 2071-2074
115. P.L. Fan and J.M. Powers, *In vitro* Wear of Aged Composite Restorative Materials, *J. Dent. Rest.,* Vol 59, 1980, p 2066-2070
116. P.L. Fan and J.M. Powers, Wear of Aged Dental Composites, *Wear,* Vol 68, 1981, p 241-248
117. J.E. McKinney and W. Wu, Chemical Softening and Wear of Dental Composites, *J. Dent. Res.,* Vol 64, 1985, p 1326-1331
118. J.-F. Roulet and C. Walti, Influence of Oral Fluid on Composite Resin and Glass-lonomer Cement, *J. Prosthet. Dent.,* Vol 52, 1984, p 182-189
119. J.C. Roberts, J.M. Powers, and R.G. Craig, Fracture Toughness of Composite and Unfilled Restorative Resins, *J. Dent. Res.,* Vol 56, 1977, p 748-753
120. A. Harrison and R.A. Draughn, Abrasive Wear, Tensile Strength and Hardness of Dental Composite Resins--Is There a Relationship?, *J. Prosthet. Dent.,* Vol 36, 1976, p 395-398
121. K.D. Jorgensen, Restorative Resins: Abrasion versus Mechanical Properties, *Scand. J. Dent. Res.,* Vol 88, 1980, p 557-568
122. A.J. De Gee, H.C. ten Harkel-Hagenaar, and C.L. Davidson, Structural and Physical Factors Affecting the Brush Wear of Dental Composites, *J. Dent. Res.,* Vol 13, 1985 p 60-70
123. A.J. De Gee, P. Pallav, and C.L. Davidson, Effect of Abrasion Medium on Wear of Stress-Bearing Composites and Amalgams *in vitro, J. Dent. Res.,* Vol 65, 1986, p 654-658
124. K.F. Leinfelder, R.W. Beaudreau, and R.B. Mazer, Device for Determining Wear Mechanisms of Posterior Composite Resins, *J. Dent. Res.,* Vol 68 (Special Issue), Abstract 330, 1989, p 908
125. J.F. Cvar and G. Ryge, Criteria for the Clinical Evaluation of Dental Restorative Materials, No. 790-244, United States Public Health Service
126. A.J. Goldberg, E. Rydinge, E.A. Santucci, and W.B. Racz, Clinical Evaluation Methods for Posterior Composite Restorations, *J. Dent. Res.,* Vol 63, 1984, p 1387-1391
127. L. Boksman, M. Suzuki, R.E. Jordan, and D.H. Charles, A Visible Light-Cured Posterior Composite Resin: Results of a 3-Year Clinical Evaluation, *J. Am. Dent. Assoc.,* Vol 112, 1986, p 627-631

128. K.F. Leinfelder, A.D. Wilder, and L.C. Teixeira, Wear Rates of Posterior Composite Resins, *J. Am. Dent. Assoc.,* Vol 112, 1986, p 829-833
129. W.F. Vann, W.W. Barkmeier, T.R. Oldenburg, and K.F. Leinfelder, Quantitative Wear Assessments for Composite Restorations in Primary Molars, *Pediatr. Dent.,* Vol 8, 1986, p 7-10
130. K.F. Leinfelder, D.F. Taylor, W.W. Barkmeier, and A.J. Goldberg, Quantitative Wear Measurements of Posterior Composite Resins, *Dent. Mater.,* Vol 2, 1986, p 198-201
131. W.F. Vann, W.W. Barkmeier, and D.B. Mahler, Assessing Composite Resin Wear in Primary Molars: Four-Year Findings, *J. Dent. Res.,* Vol 67, 1988, p 876-879
132. J.P. Moffa and A.A. Lugassy, *The M-L Scale,* Pacific Dental Research Foundation
133. S.C. Bayne, E.D. Rekow, D.F. Taylor, A.D. Wilder, J.R. Sturdevant, H.O. Heymann, T.M. Roberson, and T.B. Sluder, Laser Calibration of Leinfelder Clinical Wear Standards, *J. Dent. Res.,* Vol 69 (Special Issue), Abstract 417, 1990, p 161
134. D.F. Taylor, S.C. Bayne, J.R. Sturdevant, A.D. Wilder, and H.O. Heymann, Vivadent Comparison to M-L, Leinfelder, and USPHS Clinical Wear Scales, *J. Dent. Res.,* Vol 69 (Special Issue), Abstract 416, 1990, p 160
135. D.F. Taylor, J.R. Sturdevant, A.D. Wilder, and S.C. Bayne, Correlation of M-L, Leinfelder, and USPHS Clinical Evaluation Techniques for Wear, *J. Dent. Res.,* Vol 67 (Special Issue), Abstract 1993, 1988, p 362
136. M. Bream, P. Lambrechts, V. Van Doren, and G. Vanherle, *In vivo* Evaluation of Four Posterior Composites: Quantitative Wear Measurements and Clinical Behavior, *Dent. Mater.,* Vol 2, 1986, p 106-113
137. A.K. Abell, K.F. Leinfelder, and D.T. Turner, Microscopic Observations of the Wear of a Tooth Restorative Composite *in vivo, J. Biomed. Mater. Res.,* Vol 17, 1983, p 501-507
138. K.D. Jorgensen and E. Asmussen, Occlusal Abrasion of a Composite Restorative Resin with Ultra-Fine Filler--An Initial Study, *Quint. Int.,* Vol 9 (No. 6), 1978, p 73-78
139. J.C. Mitchem and D.G. Gronas, *In vivo* Evaluation of the Wear of Restorative Resin, *J. Am. Dent. Assoc.,* Vol 104, 1982, p 333-335
140. G.C. McDowell, T.J. Bloem, B.R Lang, and K. Asgar, *In vivo* Wear. Part I. The Michigan Computer-Graphic Measuring System, *J. Prosthet. Dent.,* Vol 60, 1988, p 112-120
141. T.J. Bloem, G.C. McDowell, B.R. Lang, and J.M. Powers, *In vivo* Wear. Part II. Wear and Abrasion of Composite Restorative Materials, *J. Prosthet. Dent.,* Vol 60, 1988, p 242-249
142. H. Xu, T. Wang, and P.A. Vingerling, A Study of Surfaces Developed on Composite Resins *in vivo* during 4-5 Years; Observations by SEM, *J. Oral Rehabil.,* Vol 16, 1989, p 407-416
143. A.D. Wilder, K.N. May, H.O. Heymann, and K.F. Leinfelder, Three Year Clinical Study of Auto-Cured Composites in Posterior Teeth, *J. Dent. Res.,* Vol 64 (Special Issue), Abstract 1604, 1985, p 353
144. A.D. Wilder, K.N. May, and K.F. Leinfelder, Three-Year Clinical Study of UV-Cured Composite Resins in Posterior Teeth, *J. Prosthet. Dent.,* Vol 50, 1983, p 26-30
145. A.D. Wilder, K.N. May, and K.F. Leinfelder, Eight Year Clinical Study of UV-Polymerized Composites in Posterior Teeth, *J. Dent. Res.,* Vol 66 (Special Issue), Abstract 481, 1987, p 167
146. L. Boksman, R.E. Jordan, M. Suzuki, D.H. Charles, and D.H. Gratton, A Five Year Clinical Evaluation of the Visible Light Cured Posterior Composite Resin Ful-fil, *J. Dent. Res.,* Vol 66 (Special Issue), Abstract 479, 1987, p 166
147. K.F. Leinfelder, Wear Patterns and Rates of Posterior Composite Resins, *Int. Dent. J.,* Vol 37, 1987, p 152-157
148. J.R. Sturdevant, T.F. Lundeen, T.B. Sluder, Jr., A.D. Wilder, and D.F. Taylor, Five-Year Study of Two Light-Cured Posterior Composite Resins, *Dent. Mater.,* Vol 4, 1988, p 105-110
149. P. Lambrechts, M. Bream, and G. Vanherle, Buonocore Memorial Lecture, Evaluation of Clinical Performance for Posterior Composite Resins and Dentin Adhesives, *Oper. Dent.,* Vol 12, 1987, p 53-78
150. F. Lutz, R.W. Phillips, J.F. Roulet, and J.C. Setcos, *In vivo* and *in vitro* Wear of Potential Posterior Composites, *J. Dent. Res.,* Vol 63, 1984, p 914-920

151. R.P. Kusy and K.F. Leinfelder, Pattern of Wear in Posterior Composite Restorations, *J. Dent. Res.,* Vol 56, 1977, p 544
152. K.-J. Soderholm, M. Zigan, M. Ragan, W. Fischlschweiger, and M. Bergman, Hydrolytic Degradation of Dental Composites, *J. Dent. Res.,* Vol 63, 1984, p 1248-1254
153. W.D. Brunson, S.C. Bayne, J.R. Sturdevant, T.M. Roberson, A.D. Wilder, and D.F. Taylor, Three-Year Clinical Evaluation of a Self-Cured Posterior Composite Resin, *Dent. Mater.,* Vol 5, 1989, p 127-132
154. K. Ratanapridakul, K.F. Leinfelder, and J. Thomas, Effect of Finishing on the *in vivo* Wear Rate of a Posterior Composite Resin, *J. Am. Dent. Assoc.,* Vol 118, 1989, p 333-335
155. R. DeLong and W.H. Douglas, A Methodology for the Measurement of Occlusal Wear, Using Computer Graphics, *J. Dent. Res.,* Vol 62, Abstract 456, 1983, p 220
156. R. DeLong, M. Pintado, and W.H. Douglas, Measurement of Change in Surface Contour by Computer Graphics, *Dent. Mater.,* Vol 1, 1985, p 27-30
157. S.C. Bayne, D.F. Taylor, J.R. Sturdevant, A.D. Wilder, H.O. Heymann, W.D. Brunson, and T.M. Roberson, Identification of Clinical Wear Factors, *J. Dent. Res.,* Vol 66 (Special Issue), Abstract 604, 1987, p 182
158. D.F. Taylor, S.C. Bayne, J.R. Sturdevant, A.D. Wilder, W.D. Brunson, and G.G. Koch, General Mathematical Model for Posterior Composite Wear, *J. Dent. Res.,* Vol 68 (Special Issue), Abstract 332, 1989, p 908
159. S.C. Bayne, H.O. Heymann, J.R. Sturdevant, A.D. Wilder, and T.B. Sluder, Contributing Co-Variables in Clinical Trials, *J. Dent. Res.,* Vol 70 (Special Issue), Abstract 14, 1991, p 267
160. I.E. Ruyter and S.A. Svendsen, Remaining Methacrylate Groups in Composite Restorative Resins, *Acta Odontol. Scand.,* Vol 36, 1978 p 75-82
161. S.I. Stupp and J. Weertman, Characterization of Monomer to Polymer Conversions in Dental Composites, *J. Dent. Res.,* Vol 58 (Special Issue), 1979, p 949
162. E. Asmussen, Restorative Resins: Hardening and Strength versus Quantity of Remaining Double Bonds, *Scand. J. Dent. Res.,* Vol 90, 1982, p 484-489
163. E. Asmussen, Factors Affecting the Quantity of Remaining Double Bonds in Restorative Resin Polymers, *Scand. J. Dent. Res.,* Vol 90, 1982, p 490-496
164. J.L. Ferracane and E.H. Greener, Fourier Transform Infrared Analysis of Degrees of Polymerization in Unfilled Resins--Methods Comparison, *J. Dent. Res.,* Vol 63, 1984, p 1093-1095
165. J.L. Ferracane, J.B. Moser, and E.H. Greener, Ultraviolet Light-Induced Yellowing of Dental Restorative Resins, *J. Prosthet. Dent.,* Vol 54, 1985, p 483-487
166. J.L. Ferracane, Correlation between Hardness and Degree of Conversion during the Setting Reaction of Unfilled Dental Restorative Resins, *Dent. Mater.,* Vol 1, 1985, p 11-14
167. J.L. Ferracane and E.H. Greener, The Effect of Resin Formulation on the Degree of Conversion and Mechanical Properties of Dental Restorative Resins, *J. Biomed. Mater. Res.,* Vol 20, 1986, p 121-131
168. A.K. Abell, K.F. Leinfelder, and D.T. Turner, *Polymer Preprints,* Vol 20, 1979, p 648-651
169. W.D. Cook, Factors Affecting the Depth of Cure of UV-Polymerized Composites, *J. Dent. Res.,* Vol 59, 1980, p 800-808
170. R. Tirtha, P.L. Fan, J.F. Dennison, and J.M. Powers, *In vivo* Depth of Cure of Photo-Activated Composites, *J. Dent. Res.,* Vol 61, 1982, p 1184-1187
171. R.L. Leung, P.L. Fan, and W.M. Johnston, Post-Irradiation Polymerization of Visible Light-Activated Composite Resin, *J. Dent. Res.,* Vol 62, 1983, p 363-365
172. E.K. Hansen, After-Polymerization of Visible Light Activated Resins: Surface Hardness versus Light Source, *Scand. J. Dent. Res.,* Vol 91, 1983, p 406-410
173. J.R. Dunn, A.H.L. Tjan, D.L. Morgan, and R.H. Miller, Curing Depths of Composites by Various Light Curing Units, *J. Dent. Res.,* Vol 66 (Special Issue), Abstract 162, 1987, p 127
174. M. Mante, V. Dhuru, and W. Brantley, Relationships between Shade, Thickness, Transmitted Curing Light Intensity and Hardness of a Composite Restorative, *J. Dent. Res.,* Vol 66 (Special Issue), Abstract 163,

1987, p 127

175. T.W. Wilson and D.T. Turner, Characterization of Polydimethacrylates and Their Composites by Dynamic Mechanical Analysis, *J. Dent. Res.,* Vol 66 (Special Issue), Abstract 167, 1987, p 127
176. V.A. Demarest and E.H. Greener, Storage Effects on Dynamic Mechanical Properties of Mica Composites, *J. Dent. Res.,* Vol 66 (Special Issue), Abstract 811, 1987, p 208
177. R.A. Draughn, Compressive Fatigue Limits of Composite Restorative Materials, *J. Dent. Res.,* Vol 58, 1979, p 1093-1096
178. C.N. Raptis, P.L. Fan, and J.M. Powers, Properties of Microfilled and Visible Light-Cured Composite Resins, *J. Am. Dent. Assoc.,* Vol 99, 1979, p 631-633
179. R. Whiting and P.H. Jacobsen, Dynamic Mechanical Properties of Resin-Based Filling Materials, *J. Dent. Res.,* Vol 59, 1980, p 55-60
180. R. Whiting and P.H. Jacobsen, A Non-Destructive Method of Evaluating the Elastic Properties of Anterior Restorative Materials, *J. Dent. Res.,* Vol 59, 1980, p 1978-1984
181. E.H. Greener, C.S. Greener, and J.B. Moser, The Hardness of Composites as a Function of Temperature, *J. Oral Rehabil.,* Vol 11, 1984, p 335-340
182. S.C. Bayne, D.F. Taylor, J.R. Sturdevant, T.M. Roberson, A.D. Wilder, and M.W. Lisk, Protection Theory for Composite Wear Based on 5-Year Clinical Results, *J. Dent. Res.,* Vol 67 (Special Issue), Abstract 60, 1988, p 120
183. J.C. Roberts, J.M. Powers, and R.G. Craig, Wear of Commercial Pit and Fissure Sealants, *J. Dent. Res.,* Vol 56, 1977, p 692
184. M. Houpt, Z. Shey, A. Chosack, E. Eidelman, A. Fuks, and J. Shapira, Occlusal Composite Restorations: 4-Year Results, *J. Am. Dent. Assoc.,* Vol 110, 1985, p 351-353
185. J.M. Powers, J.A. Capp, and R.G. Craig, Abrasion of Temporary Filling Materials, *J. Mich. Dent. Assoc.,* Vol 56, 1974, p 281-283
186. J.M. Powers, P.L. Fan, and R.W. Hostetler, Properties of Class V Restorative Materials, *J. Mich. Dent. Assoc.,* Vol 63, 1981, p 275-278
187. J.E. McKinney, J.M. Antonucci, and N.W. Rupp, Wear and Microhardness of a Silver-Sintered Glass-Ionomer Cement, *J. Dent. Res.,* Vol 67, 1988, p 831-835
188. G.T. Charbeneau and R.R. Bozell III, Clinical Evaluation of a Glass Ionomer Cement for Restoration of Cervical Erosion, *J. Am. Dent. Assoc.,* Vol 98, 1979, p 936-939
189. A. Ekfeldt and G. Oilo, Wear of Prosthodontic Materials--An *in vivo* Study, *J. Oral Rehabil.,* Vol 17, 1990, p 1-13
190. J.R. Beall, Wear of Acrylic Resin Teeth, *J. Am. Dent. Assoc.,* Vol 30, 1943, p 252-256
191. J.A. Saffir, Further Studies in Evaluating Acrylics for Masticatory Surface Restorations, *J. Am. Dent. Assoc.,* Vol 31, 1944, p 518-523
192. J. Osborne, Abrasion Resistance of Dental Materials, *Brit. Dent. J.,* Vol 87, 1949, p 10-12
193. F.A. Slack, A Preliminary Method for Testing Abrasion Hardness, *J. Am. Dent. Assoc.,* Vol 39, 1949, p 47-50
194. G.E. Monasky and D.F. Taylor, Studies on Wear of Porcelain, Enamel, and Gold, *J. Prosthet. Dent.,* Vol 25, 1971, p 299-306
195. C.N. Raptis, J.M. Powers, and P.L. Fan, Frictional Behavior and Surface Failure of Acrylic Denture Teeth, *J. Dent. Res.,* Vol 60, 1981, p 908-913
196. C.N. Raptis, J.M. Powers, and P.L. Fan, Wear Characteristics of Porcelain Denture Teeth, *Wear,* Vol 67, 1981, p 177-185
197. A. Koran, R.G. Craig, and E.W. Tillitson, Coefficient of Friction of Prosthetic Tooth Materials, *J. Prosthet. Dent.,* Vol 27, 1972, p 269-274
198. V.W. Boddicker, Abrasion Tests for Artificial Teeth, *J. Am. Dent. Assoc.,* Vol 35, 1947, p 793-797
199. J.A. Mahalik, F.J. Knap, and E.J. Weiter, Occlusal Wear in Prosthodontics, *J. Am. Dent. Assoc.,* Vol 82, 1971, p 154-159

200. A.R.T. Greenwood, Wear Testing Equipment for Synthetic Resin Teeth, *Abstr. J. Dent. Res.,* Vol 34, 1955, p 741-742

201. J.C. Thompson, Attrition of Acrylic Teeth, *Dent. Pract. (Bristol),* Vol 15, 1965, p 233-236

202. R.E. Myerson, The Use of Porcelain and Plastic Teeth in Opposing Complete Dentures, *J. Prosthet. Dent.,* Vol 7, 1957, p 625-633

203. J.A. Cornell, J.S. Jordan, S. Ellis, and E.E. Rose, A Method of Comparing the Wear Resistance of Various Materials Used for Artificial Teeth, *J. Am. Dent. Assoc.,* Vol 54, 1957, p 608-614

204. J.P. Coffey, R.J. Goodkind, R. DeLong, and W.H. Douglas, *In vitro* Study of the Wear Characteristics of Natural and Artificial Teeth, *J. Prosthet. Dent.,* Vol 54, 1985, p 273-280

205. W.R. Mann and O.C. Applegate, Acrylic Teeth as a Means of Obtaining Balanced Functional Occlusion in Partial Denture Prosthesis, *J. Am. Dent. Assoc.,* Vol 31, 1944, p 505-514

206. P.G. Lofberg, Pronounced Abrasion of Acrylic Teeth, *Dent. Abstr.,* Vol 2, 1957, p 432-433

207. A.S.T. Franks, Clinical Appraisal of Acrylic Tooth Wear, *Dent. Pract. (Bristol),* Vol 12, 1962, p 149-153

208. V.H. Sears, Occluding Porcelain to Resin Teeth-A Clinical Evaluation, *Dent. Surv.,* Vol 36, 1960, p 1144-1146

209. A. Ekfeldt and G. Oilo, Occlusal Contact Wear of Prosthodontic Materials an *in vivo* Study, *Acta Odont. Scand.,* Vol 46, 1988, p 159-169

210. A. Ekfeldt and G. Oilo, Wear Mechanisms of Resin and Porcelain Denture Teeth, *Acta Odont. Scand.,* Vol 47, 1989, p 391-399

211. A. Harrison, Clinical Results of the Measurement of Occlusal Wear of Complete Dentures, *J. Prosthet. Dent.,* Vol 35, 1976, p 504-511

212. R.E. Ogle, L.J. David, and H.R. Ortman, Clinical Wear Study of a New Tooth Material. II, *J. Prosthet. Dent.,* Vol 54, 1985, p 67-75

213. L.P. Adams, C.H. Jooste, and C.J. Thomas, An Indirect *in vivo* Method for Quantification of Wear of Denture Teeth, *Dent. Mater.,* Vol 5, 1989, p 31-34

214. A. Harrison, R. Huggett, and R.W. Handley, A Correlation between Abrasion Resistance and Other Properties of Some Acrylic Resins Used in Dentistry, *J. Biomed. Mater. Res.,* Vol 13, 1979, p 23-34

215. R. Lappalainen, A. Yli-Urpo, and L. Seppa, Wear of Dental Restorative and Prosthetic Materials *in vitro, Dent. Mater.,* Vol 5, 1989, p 35-37

216. G.R. Miller, J.M. Powers, and K.C. Ludema, Frictional Behavior and Surface Failure of Dental Feldspathic Porcelain, *Wear,* Vol 31, 1975, p 307-316

217. R. DeLong, W.H. Douglas, R.L. Sakaguchi, and M.R. Pintado, The Wear of Dental Porcelain in an Artificial Mouth, *Dent. Mater.,* Vol 2, 1986, p 214-219

218. P.L. Fan, J.M. Powers, and B.C. Reid, Surface Mechanical Properties of Stone, Resin, and Metal Dies. *J. Am. Dent. Assoc.,* Vol 103, 1981, p 408-411

219. Y. Shoji, Studies of the Mechanism of the Mechanical Enlargement of Root Canals, *Nihon Univ. Sch. Dent. J.,* Vol 7, 1965, p 71-78

220. H. Ochia, Studies on Dental Hand Reamer. I. Automatic Measurement of the Reamer Diameter and Cutting Torque; II. Diameter and Cutting Sharpness of Commercial Reamers, *Jpn. J. Conserv. Dent.,* Vol 19, 1976, p 41-73

221. S. Oliet and S.M. Sorin, Cutting Efficiency of Endodontic Reamers, *Oral Surg., Oral Path., Oral Med.,* Vol 36, 1973, p 243-252

222. J. Weber, J.B. Moser, and M.A. Heuer, A Method to Determine the Cutting Efficiency of Root Canal Instruments in Linear Motion, *J. Endodont.,* Vol 6, 1980, p 829-834

223. J.G. Newman, W.A. Brantley, and H. Gerstein, A Study of the Cutting Efficiency of Seven Brands of Endodontic Files in Linear Motion, *J. Endodont.,* Vol 9, 1983, p 316-322

224. R.G. Neal, R.G. Craig, and J.M. Powers, Cutting Ability of K-Type Endodontic Files, *J. Endodont.,* Vol 9, 1983, p 52-57

225. E. Stenman and L.S.W. Spangberg, Machining Efficiency of Endodontic Files: A New Methodology, *J.*

Endodont., Vol 16, 1990, p 151-157

226. R.G. Neal, R.G. Craig, and J.M. Powers, Effect of Sterilization and Irrigants on the Cutting Ability of Stainless Steel Files, *J. Endodont.,* Vol 9, 1983, p 93-96
227. R.A. Felt, J.B. Moser, and M.A. Heuer, Flute Design of Endodontic Instruments: Its Influence on Cutting Efficiency, *J. Endodont.,* Vol 8, 1982, p 253-259
228. L.J. Miserendino, J.B. Moser, M.A. Heuer, and E.M. Osetek, Cutting Efficiency of Endodontic Instruments. Part II: Analysis of Tip Design, *J. Endodont.,* Vol 12, 1986, p 8-12
229. J. Lindhe and L. Jacobson, Evaluation of Periodontal Scalers. I. Wear Following Clinical Use, *Odontol. Revy,* Vol 17, 1966, p 1-8
230. J. Lindhe, Evaluation of Periodontal Scalers. II. Wear Following Standardized Orthogonal Cutting Tests, *Odontol. Revy,* Vol 17, 1966, p 121-130
231. J. Lindhe, Evaluation of Periodontal Scalers. III. Orthogonal Cutting Analyses, *Odontol. Revy,* Vol 17, 1966, p 251-273
232. H. Tal, A. Kozlovsky, E. Green, and M. Gabbay, Scanning Electron Microscope Evaluation of Wear of Stainless Steel and High Carbon Steel Curettes, *J. Periodontol.,* Vol 60, 1989, p 320-324
233. M.P. Benfenati, M.T. Montesani, S.P. Benfenati, and D. Nathanson, Scanning Electron Microscope: An SEM Study of Periodontally Instrumented Root Surfaces, Comparing Sharp, Dull and Damaged Curettes and Ultrasonic Instruments, *Int. J. Periodont. Rest. Dent.,* Vol 2, 1987, p 51-67
234. H. Tal, J.M. Panno, and T.K. Vaidyanathan, Scanning Electron Microscope Evaluation of Wear of Dental Curettes during Standardized Root Planing, *J. Periodontol.,* Vol 56, 1985, p 532-536
235. D.C. Tidy, Frictional Forces in Fixed Appliances, *Am. J. Orthod. Dentofac. Orthop.,* Vol 96, 1989, p 249-254
236. J. Nicolls, Frictional Forces in Fixed Orthodontic Appliances, *Dent. Prac. Dent. Rec.,* Vol 18, 1967-1968, p 362-366
237. G.F. Andreasen and F.R. Quevedo, Evaluation of Friction Forces in the 0.022 × 0.028 Edgewise Bracket *in vitro, J. Biomech.,* Vol 3, 1970, p 151-160
238. C.A. Frank and R.J. Nikolai, A Comparative Study of Frictional Resistance between Orthodontic Bracket and Arch Wire, *Am. J. Orthod.,* Vol 78, 1980, p 593-609
239. J.G. Stannard, J.M. Gau, and M.A. Hanna, Comparative Friction of Orthodontic Wires under Dry and Wet Conditions, *Am. J. Orthod.,* Vol 89, 1986, p 485-491
240. L.D. Garner, W.W. Allai, and B.K. Moore, A Comparison of Frictional Forces during Simulated Canine Retraction of a Continuous Edgewise Bracket, *Am. J. Orthod. Dentofac. Orthop.,* Vol 90, 1986, p 199-203
241. L. Peterson, R. Spencer, and G. Andreasen, Comparison of Frictional Resistance for Nitinol and Stainless Steel Wire in Edgewise Arch Wire, *Quint. Int.,* Vol 13 (No. 5), 1982, p 563-571
242. D. Drescher, C. Bourauel, and H.-A. Schumacher, Frictional Forces between Bracket and Arch Wire, *Am. J. Orthod. Dentofac. Orthop.,* Vol 96, 1989, p 397-404

Friction and Wear of Electrical Contacts

Ralph A. Burton, Burton Technologies Inc.

Introduction

ELECTRICAL BRUSHES running on metallic slip rings or commutators represent one of the oldest applications of dry sliding contacts. Brush performance is influenced by parameters such as current, sliding speed, load, contact dynamics, and ambient atmosphere. Although a carbon brush gliding over a copper current collector (slip ring or commutator) was successfully employed in electrical power machinery over 100 years ago, research actively continues on:

- Improved composite materials for higher conductivity and lower wear
- Low electrical noise in instrumentation
- Materials resistant to arcing
- Substitute materials to replace compositions that are hazardous in manufacture or use

A separate field of development has focused on the use of static contacts for applications such as communications and computer systems where noise, development of contaminant films, fretting under vibration, and resistance to attack by atmospheric components are concerns.

Research contributions have come from many disciplines, and the literature published annually is extensive. A continuing source of timely literature is the annual Institute of Electrical and Electronics Engineers (IEEE) Holm Conference on Electrical Contacts, sponsored by the Components, Hybrids, and Manufacturing Technology Society of the Institute of Electrical Engineers, Inc.*

The classic textbook in the field of electrical contacts was written by Holm (Ref 1). It is now out of print but is must reading if it can be located.

A later textbook, also out of print, is by Shobert (Ref 2). A relatively recent summary on carbon materials, which includes glassy carbon, is given by Kinoshita (Ref 3).

A 1982 issue of *Wear* (Ref 4) contains 25 papers that collectively provide a summary of the electrical contact field from the viewpoint of tribology. This publication includes papers on fiber brushes and very high performance brushes.

Seventy-five papers are featured in a publication edited by Ji-Gao Zhang (Ref 5). It includes a concise yet very informative summary, written by Morton Antler, of the tribology of contact finishes for electronic connectors.

Note

* *Copies of the proceedings may be obtained through the IEEE Service Center, 445 Hoes Lane, P.O. Box 1331, Piscataway, N.J. 08855-1331. Many of the papers ultimately appear in the IEEE Proceedings. A cumulative index through 1989 has been prepared (W.L. Brodsky, Task Group Chairman) and is available as IEEE Catalog Number JH 9412-H.

Contact Resistance

The passage of current through a sliding contact of solids is accompanied by a contact potential drop (voltage), in addition to the drop through the bodies themselves. This drop is exacerbated by the fact that solids only contact one another at the peaks of roughness asperities. The actual contact area is determined by the hardness, P_H, of the softer material of the contact pair. Because normal stress on brushes and sliders tends to be low in order to minimize frictional dissipation, the actual area of contact, A_c, may be as small as 10^{-5} A_N, where A_N is the nominal contact area (area of the brush face).

For an isolated circular contact spot, Holm (Ref 1) has derived contact resistance as:

$$R = r/2a \qquad \textbf{(Eq 1)}$$

where R is the resistance in Ω, r is the resistivity in $\Omega \cdot$ cm, and a is the radius of the contact spot.

For n independent contact spots conducting in parallel, the combined resistance, R_N, is given by:

$$R_N = r/2an \qquad \textbf{(Eq 2)}$$

If, however, the spots are clustered over a patch of radius, α, Greenwood (Ref 6) has shown:

$$R_N = (r/2)\,(1/na + 1/\alpha) \qquad \textbf{(Eq 3)}$$

This can be written as (Ref 7):

$$R_N = (r/2\alpha)\,[1 + (A_N/nA_C)^{1/2}] \qquad \textbf{(Eq 4)}$$

Williamson and Hunt (Ref 8) have shown that for high contact loads (for example, between two spherical surfaces), the ratio A_N/A_c is $\sim$0.5 even when the nominal normal stress on the contact patch, P_N, exceeds the Brinell pressure, P_H, under which the surfaces take a permanent indentation. In that case, for any reasonable n (for example, n = 200), the second term in brackets of Eq 4 becomes 0.1, and is therefore small relative to the first term. The resistance, then, is that of the patch of radius α as if it were in full contact.

This observation explains why the smaller asperities, or subasperities, on larger asperities need not affect conduction. It also explains why resistance measurements (both thermal and electric) indicate a few relatively large contact patches on a sliding brush rather than large numbers of microscopic ones.

Effect of Insulating Films (Fritting). The above concepts can provide a guide to estimate the contact resistance of untarnished noble metals in contact. However, even the simplest system (that is, carbon on copper) develops an insulating film when operating in air. Early experiments showed this film to have an amorphous structure and to be composed of copper oxide and carbon. Similar to oxide films in other applications, such a film is prone to breakdown when the potential gradient at asperity junctions exceeds a limiting value. When a breakdown at one location leads to formation of a conductive junction, this reduces the potential difference between the carbon and the copper. This phenomenon is known as fritting.

When a carbon-copper contact is placed in series with resistances, and potential across the assembly is increased, contact potential rises to a characteristic magnitude and remains at this magnitude as current is increased. This happens because an increase of potential across nonfritted junctions results when increased current passes through those junctions that have been fritted. Increased potential leads to increased numbers of fritted junctions.

Figure 1 shows data compiled by Shobert (Ref 2) for composite brushes that contained a varied metal content. The fritting potential is dependent upon the polarity of the brush and is typically lower for the cathodic contacts. In terms of diffusion mechanics, it has been theorized that the cathodic potential encourages copper ions to diffuse from the substrate into the film, thus reducing its resistance and encouraging breakdown of the film. This mechanism is aided by local defects in the insulating film. The initial breakdown and formation of a metallic pipe through the film is called A-fritting; the growth of conductivity with increased current, known as B-fritting, is hypothesized to proceed by a different mechanism that involves resistance heating of the conductive pipes.

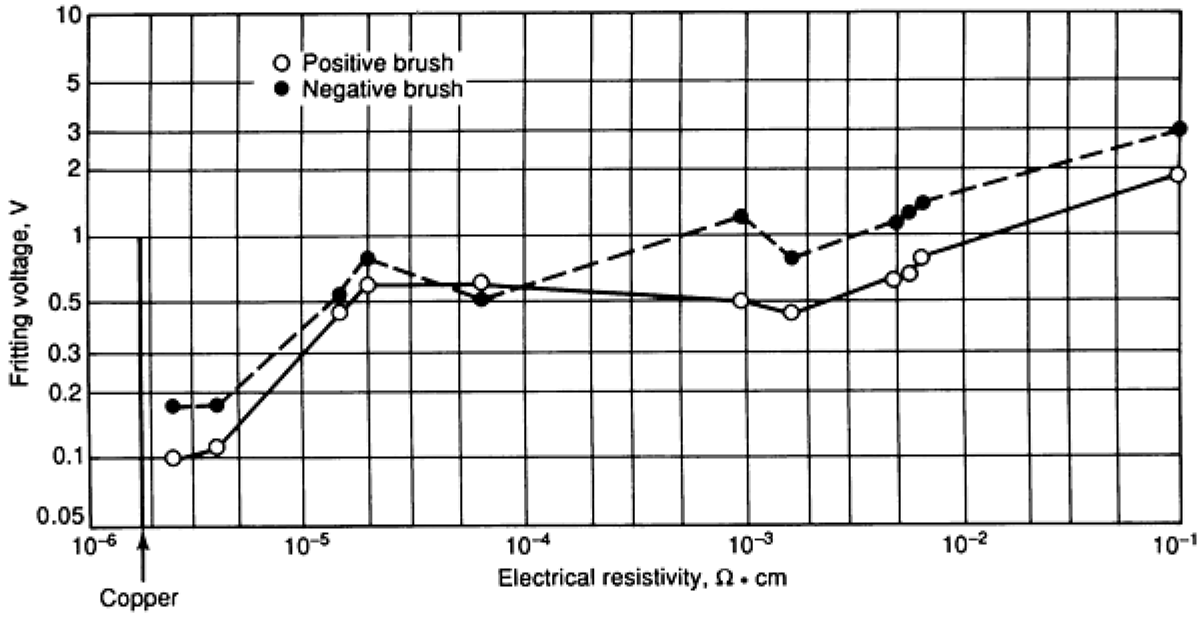


Fig. 1 Plot of fritting voltage versus electrical resistivity showing polarity dependence of composite brushes. Voltage was generated by the contact potential drop between a copper collector and a carbon/copper composite brush separated by an insulating film that was produced by rubbing action of the collector and the brush components.

Resistivity of the film is not as important a factor in thermal conduction as it is in electrical conduction because the ratio of thermal resistivities for the insulator and the metal is several orders of magnitude smaller than the ratio of electrical resistivities.

Tunneling. When thin insulating films are used between two conductors, electrons can freely move across the films thicknesses of $\leq$2 nm ($\leq$20 Å). In Fig. 2, Shobert shows how tunneling resistance varies with film thickness and work function, Φ, of the conductors. The shaded region represents the domain of a typical brush (for example, carbon on copper) where the work function, Φ, is between expected values of 2 and 4 V. In the sliding brush, the fritted spots may oxidize between contacts. However, if film growth is kept to <1 nm (<10 Å) because of wear, tunneling will occur with very little resistance.

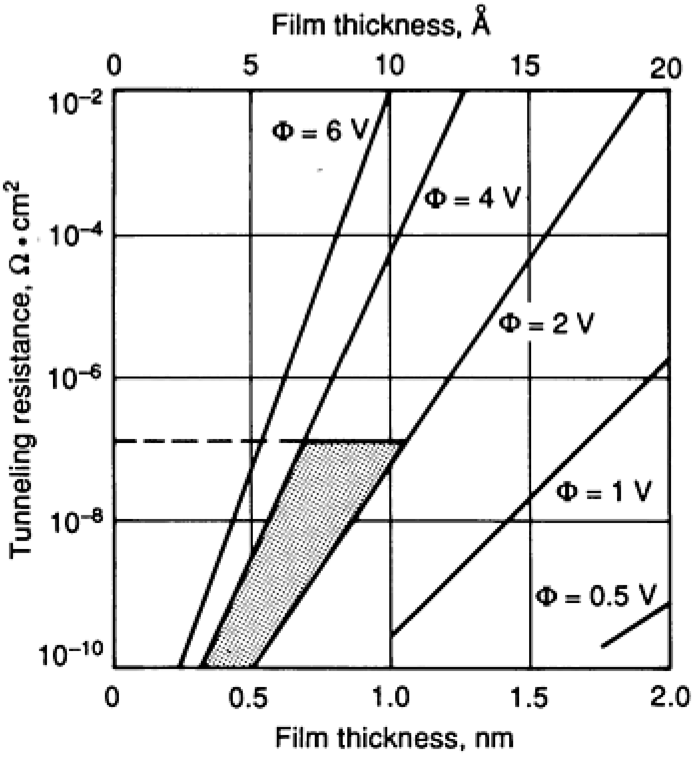


Fig. 2 Plot of tunneling resistance versus film thickness at selected work function, Φ, values. Shaded region represents typical operating range for carbon-on-copper brush components.

Mechanical Factors. Reichner (Ref 9) has presented an overview of the effects of geometric errors in a current collector. In steady operation, the brush wears in to conform to the radius of the current collector. However, both angular and radial runout tend to cause the brush face to develop a relative rounding as though part of a toroidal surface. The contact patch eventually becomes elliptical.

For simple spherical contact, the radius α of the contact patch is approximately:

$$\alpha = 0.880(F_N\Gamma/E)^{1/3} \qquad \textbf{(Eq 5)}$$

where E is Young's modulus for the carbon (E value for the metal is typically high and can be neglected). Poisson's ratio has been taken as 0.3, and $1/\Gamma$ is a composite of the principal relative curvatures of the contacting bodies:

$$1/\Gamma = (1/\Gamma_1 + 1/\Gamma_2) \qquad \textbf{(Eq 6)}$$

Thermoelastic Mounding. A brush and a collector that conform to one another in the static condition may be subject to thermoelastic deformation, which leads to a quasispherical mounding and a quasi-Hertzian contact patch formation. In contrast to the simple elastic spherical deformation discussed in the section "Mechanical Factors" of this article, thermoelastic mounding does not necessarily respond to an increase in load with an increase in the α value.

Frictional Heating. When frictional heating predominates, an increase in the load results in an increase in the thermal deformation component, which causes a corresponding decrease in Γ. The net effect is to make α almost independent of load. In that case, α would be primarily dependent upon sliding speed.

Electrical Heating. When electrical heating is the dominant effect, an increase in load will reduce contact resistance and, upon consequent heating, will result in an increase in Γ.

Heat Transfer onto Slip Ring. For typical slip ring geometry, both the frictional heat and electrical heat generated are transferred with negligible heat loss into the metallic ring, which subsequently develops a mound that is nearly stationary in space at the contact spot. Because of the low value of Young's modulus for typical carbon brushes, mounding of the brush is not a major factor. Experiments (Ref 10) indicate that this phenomenon can occur at modest sliding speed, leading to $\alpha \leq 1$ mm (0.04 in.).

Burnout. Williamson (Ref 11) has analyzed the thermal problem of resistive heating of a stationary contact spot in which the electrical resistivity rises with temperature. An instability exists at which point the rising resistance leads to runaway heating. Similar to the fuse-wire burnout phenomenon that occurs at a critical current level, a fuse wire will begin an accelerated rise in temperature until it burns out. It is also a process that is exploited in the spot welding of metallic sheets or plates. This phenomenon could possibly be involved in A-fritting. Bryant (Ref 12) recently published an analysis of thermal, stress, and electrical behavior of a three-dimensional model of this phenomenon.

Circuit Breaking

Circuit breaking under current flow is a primary wear phenomenon in switches and is an important factor in direct current (dc) vehicular electric systems. Although electrical motors are designed to minimize current flow at the moment of breaking in commutation, residual effects are always present. Inadvertent breaking occurs when runout, which is coupled with inadequate brush-holder design, leads to bouncing against a current collector.

In a gaseous atmosphere, the breaker points may be paralleled by a capacitance that accepts the displaced current and provides a controlled potential rise between the breaker points. If the rate of separation is rapid enough, the potential gradient in the gas between the points can be kept from exceeding the breakdown potential of the gas. Should a brief plasma discharge occur, refractory metal points or hard carbons can resist the heat-flash with minimal damage.

The mechanical components of circuit breakers for energy distribution systems are designed to allow the points to acquire a relative velocity before parting. In such systems, acceleration from a stationary contact occurs in sliding, and breaking occurs when one point reaches the edge of the other.

Dow and Kannel (Ref 13) have carried out experiments with a gravity-loaded brush in which bouncing took place to show the effects of coupled thermoelastic wear and electrical discharge.

Bryant's model (Ref 12) accounts for some of the surface damage. For brushes that have normal oscillations relative to the current collector, the Williamson effect (Ref 11) causes disintegration of a packet of solid material as the conduction is funneled through spots of decreasing size or diminishing quantity.

Takagi (Ref 14) has written a summary of his own work and that of others on the process of contact breaking. The effect of material migration on contact breaking is shown in Fig. 3.

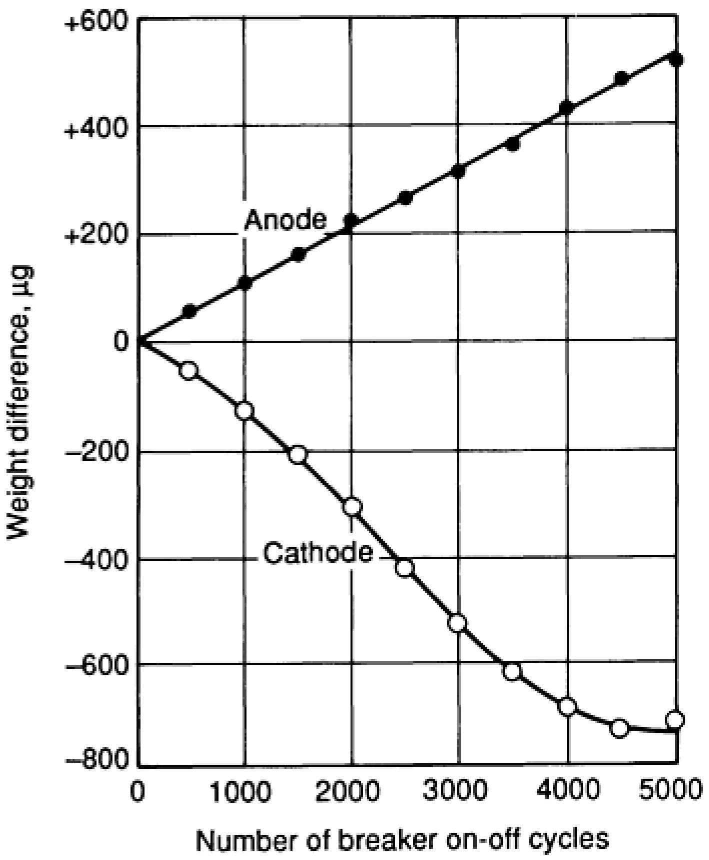


Fig. 3 Metal transfer (weight gain or loss) between silver contact breaker points at a current of 3 A

Atmospheric Effects

Graphitic carbons, in particular, depend upon atmospheric components such as water to weaken the bond between layers, to promote sliding, and to provide a bond between graphite particles and the metallic oxide substrate. In dry (dew point, <10 °C, or 50 °F) or inert gas environments, the phenomenon of dusting, whereby rapid wear of the sliding brush produces copious amounts of carbon dust, is encountered.

Dusting can be controlled with the addition of molybdenum disulfide, barium fluoride, or lithium carbonate into the brush microstructure (Ref 2). In addition, alcohols or even aliphatic hydrocarbons in the atmosphere can minimize the dusting problem. Air is the atmosphere of choice due to its availability; however, moist hydrogen is preferred when available because it will cause one-tenth of the wear rate that is generated in air. This is also true of other inert atmospheres such as ammonia and carbon dioxide (Ref 15). Recent studies of glassy carbon indicate that this form of carbon is immune to dusting even without additives in dry nitrogen (Ref 16).

Greatly accelerated wear can be caused by silicone vapor (Ref 17) as low as 10 ppm in ambient air. This is moderated by the addition of cadmium fluoride, which converts silica to a gaseous material. Solvents and paint thinners may also accelerate wear.

Brush Materials

Traditional carbon brushes are produced with the aid of organic binders, which are used to blend and form powders of lampblack, synthetic or natural graphites, amorphous carbons, dusting inhibitors, and metallic powders. These compositions are then fired slowly at temperatures >1000 °C (>1830 °F). A tremendous range of hardness and resistivity results from the propriety compositions of materials used in brushes. Determine the materials used for various applications is arrived at through experience with similar applications. Brush "reading," or damage inspection, can lead to recommended changes in hardness, transfer film reaction, or metallic content.

Initial evaluation of composites formed by massive electrical discharges (Ref 18) and composites formed from glassy carbon (Ref 16) indicates that these carbon-based materials have potential uses as brush materials.

References

1. R. Holm, *Electrical Contacts Handbook*, Springer-Verlag, Berlin, 1958
2. E. Shobert, *Carbon Brushes,* Chemical Publishing, 1965
3. K. Kinoshita, *Carbon,* John Wiley & Sons, 1988
4. *Wear,* Vol 78, 1982
5. Ji-Gao Zhang, Ed., *Electrical Contacts and Electromechemical Components,* International Academic Publishers, Pergamon Press, Oxford, 1989
6. J.A. Greenwood, Constriction Resistance and the Real Area of Contact, *Brit. J. Appl. Phys.*, Vol 17, 1966, p 1621
7. R.A. Burton and R.G. Burton, Cooperative Interactions of Asperities in the Thermotribology of Sliding Contact, *Proceedings of the IEEE Holm Conference,* Institute of Electrical and Electronics Engineers, 1990, p 256
8. J.B.P. Williamson and R.T. Hunt, Asperity Persistence and the Real Area of Contact Between Rough Surfaces, *Proc. R. Soc. (London) A,* Vol 327, 1972, p 147
9. P. Reichner, Wear Induced Profiles of Brushes on Eccentric Rotors, *Wear,* Vol 78, 1982, p 207
10. R.A. Burton and R.G. Burton, Experiments and Analysis of the Thermoelastic Behavior of a Brush on a Slipring, *Proceedings of the IEEE Holm Conference,* Institute of Electrical and Electronic Engineers, 1989, p 25
11. J.P.B. Williamson and N. Allen, Thermal Stability in Graphite Contacts, *Wear,* Vol 78, 1982, p 39
12. M.D. Bryant, A. Particle Ejection Mechanism for Brushwear, *Proceedings of the IEEE Holm Conference,* Institute of Electrical and Electronics Engineers, 1990, p 285
13. T.A. Dow and J.W. Kannel, Thermomechanical Effects in High Current Density Electrical Slip Rings, *Wear,* Vol 79, 1982, p 93
14. T. Takagi, Relationship Between Contact Operating Conditions and Contact Phenomena in Ag and Pd Breaking Contacts, *Proceedings of the IEEE Holm Conference,* Institute of Electrical and Electronics Engineers, 1990, p 1
15. R.G. Burton and R.A. Burton, Tribology of Carbon Matrix Composites, *Tribology of Composites Materials,* P. Rohatgi, P. Blau, and C. Yust, Ed., ASM International, 1990, p 309
16. J. Marsden and R.H. Savage, Effects of Silicone Vapor on Brush Wear, *Electr. Eng.,* Vol 67 (No. 1084), 1948,p 174
17. W.G. Krellner, "Silicon Insulated Dynamoelectric Machine," U.S. patent 2,855,528, 1958
18. C. Persad, S.C. Sparks, D. Moore, M. Schmerling, Z. Eliezer, J. Gully, and R. Carnes, Advanced Composite Materials for High-Performance Electrobiological Applications, *Tribology of Composite Materials,* P. Rohatgi, P. Blau, and C. Yust, Ed., ASM International, 1990, p 203

Friction and Wear of Semiconductors

Steven Danyluk and Soo-Wohn Lee, University of Illinois at Chicago

Introduction

SEMICONDUCTOR CIRCUITS are produced using a variety of mechanical deformation processes. Single-crystal semiconductors, such as silicon and gallium arsenide, are grown from the melt into solid cylinders (ingots) that must be shaped, polished, and cut into geometries suitable for handling and processing (Ref 1). Figure 1 illustrates the typical processing sequence for the production of waters that serve as substrates for large-scale integrated (LSI) circuits. The diameter of the as-grown ingot is usually not uniform. Therefore, after the ends are cropped, the outside surface is abrasively ground to produce a precise diameter. In addition, an orientation flat is ground along the length of the cylinder to indicate the dopant type and the crystallographic orientation. The ingot is etched to remove the gross grinding damage. Then sliced by diamond-impregnated blades into wafers approximately 1 mm (0.04 in.) thick. The wafer circumference is made uniform by grinding along the perimeter, and the faces are lapped with an abrasive slurry. The lapped wafers are etched and then polished on one side.

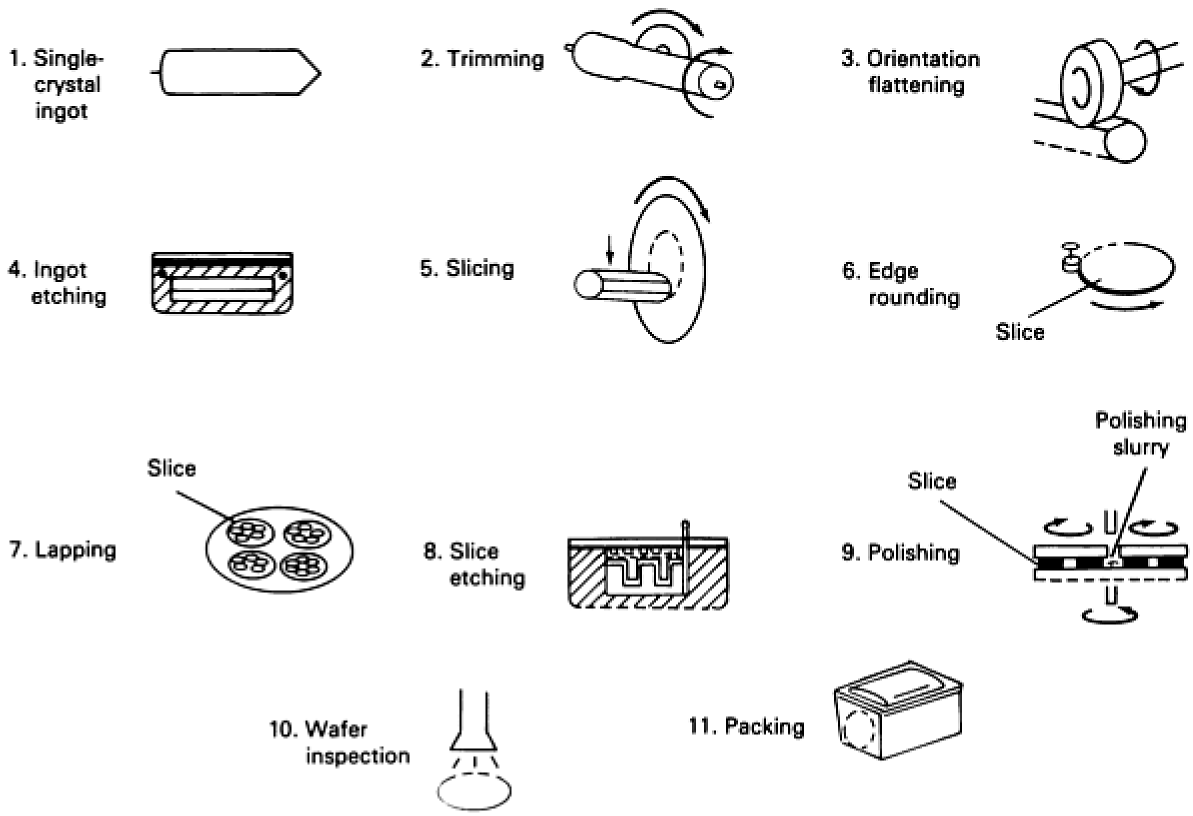


Fig. 1 Typical processing sequence required to produce wafers that serve as substrates for LSI circuits

LSI circuits are fabricated on the polished side of a wafer by a variety of diffusion, deposition, and heat treatment steps. The complexity of a particular device, or die, determines the number of steps needed to complete the circuit, as well as the physical dimensions of the circuit. The dies are removed by dicing the wafer with diamond-impregnated blades. They are then mounted on a substrate and electrical leads are attached by the physical positioning of metal wires onto thin-film metal pads.

Sawing, grinding, lapping, polishing, and dicing involve frictional contact between abrasive particles and the semiconductor surfaces, resulting in surface microcracks and dislocations. The depth to which these defects propagate beneath the surface depends on the contact stresses, temperatures, and lubricants used in the processing, and on the doping level and orientation of the semiconductor. Microcracks are particularly insidious in single-crystal semiconductors because of the brittle nature and anisotropic mechanical properties of these materials (Ref 2).

Microcracks propagate on cleavage planes at very low stresses (Ref 3), and dislocations can glide on slip planes over large distances beyond the immediate contact zone of the abrasive particles (Ref 4). Thus, precise control of the size of the damage zone produced by grinding or dicing, and elimination of cracking by polishing or etching are impossible unless the cutting variables are controlled and the deformation mechanisms are known. Failure to remove microcracks

may result in unexpected catastrophic failure of wafers or dies, and failure to remove dislocations may degrade the electrical properties of the dies (Ref 5). In either case, the yield of usable devices will be decreased.

Because friction and wear are important in the processing of semiconductor devices, much effort has gone into determination of the mechanisms of the surface deformation caused by frictional contact. Research hsa focused on the friction of abrasive particles in contact with semiconductor surfaces, the chemical interaction with lubricants, and the influence of dopants and crystallographic orientation. This article describes the result of several of these studies.

Mechanical Damage at Silicon Surfaces Caused by Wafering

Wafering of ingots is accomplished by string, reciprocating blade, or inner diameter (ID) sawing (Fig. 2). These three methods (Ref 6) rely on the abrasive action or hard particles for comminution of the semiconductor in the vicinity of the contact of the abrasive particles. For example, string and blade saws work by carrying an abrasive slurry over the surface of the ingot. The abrasive particles contact the surface and rotate and tumble as the string or blade moves in either a straight-through or a reciprocating motion (Fig. 2a). The fluid carries the abrasive and removes and the frictional heat generated in the process.

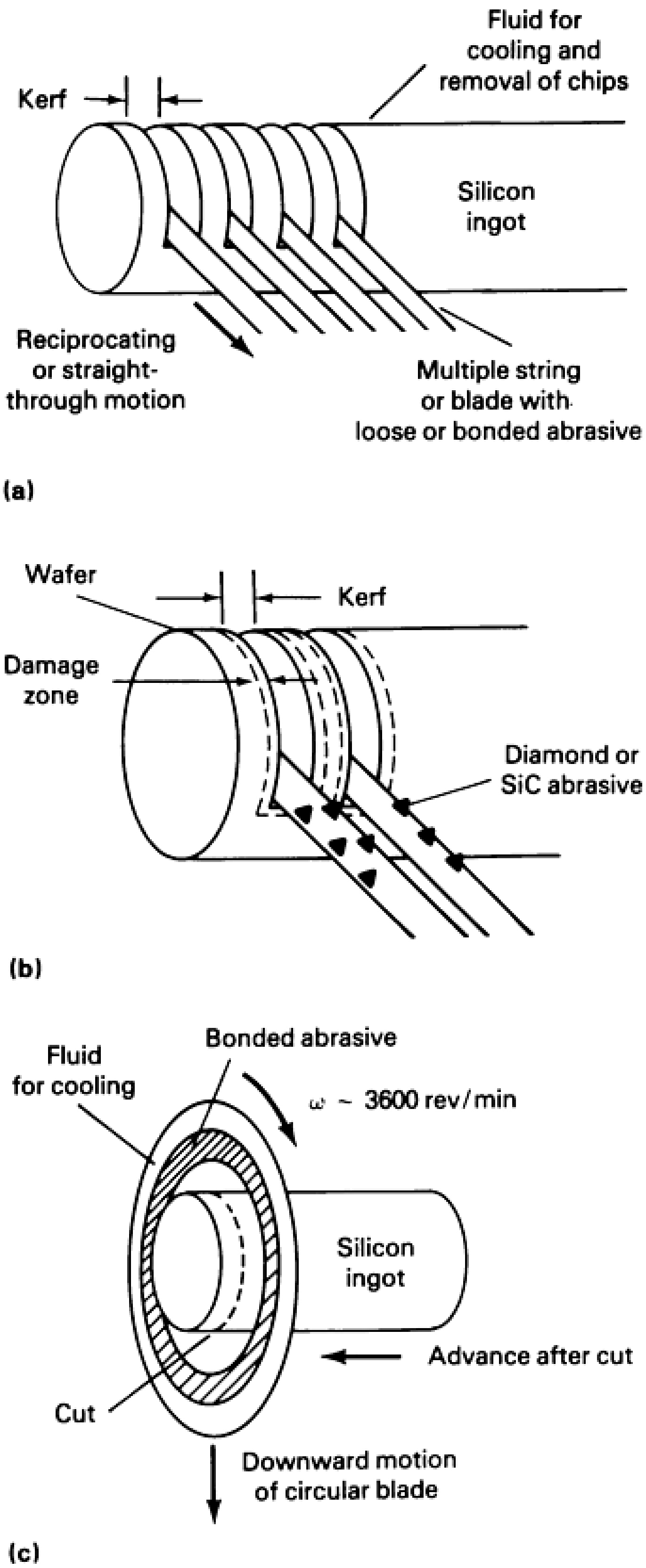


Fig. 2 Processing of ingots into wafers using two types of cutting techniques. (a) Typical multiple string or reciprocating blade sawing setup for wafering. (b) Closeup of damage zone generated by string or blade kerf. (c) Inner diameter sawing

ID sawing is a special type of blade sawing in which the abrasive is bonded to a circular wheel that rotates at high speed (Fig. 2b). Because of the slice-to-slice thickness uniformity and quality, approximately 80% of wafers are produced by ID

sawing (Ref 6). In this process, a thin steel blade (0.3 to 0.4 mm, or 0.01 to 0.02 in., thick) with a 0.10 to 0.15 mm (0.004 to 0.006 in.) diam hole in the center is tensioned in a fixture. The ingot is inserted in the central hole as the lubricant-cooled blade is rotated at 3600 rev/min. The inner surface of the hole is coated with an electrodeposited nickel film that contains a 0.5 μm diamond grit. The contact of the diamond with the silicon creates cracks and dislocations. If the cracks intersect the surface, conchoidal fracture will produce comminution of the silicon. The width of the cut, called the kerf, depends on the blade parameters (thickness, vibration, and so on), dopant type and crystallographic orientation of the ingot, and lubricant properties. A typical as-cut wafer thickness is 1.0 mm (0.04 in.). Other requirements include parallelism within 0.03 mm (0.0012 in.), flatness within +0.03 mm (0.0012 in.), and surface damage (depth) of less than 0.025 mm (0.0010) in.).

Much research has focused on the damage caused by cropping (removal of unusable sections of the ingot), centerless grinding (mill grinding to produce uniform shapes), and exit chipping (fracture of wafers and ingots as the blade exits the cut) (Ref 7), the influence of lubricants on the dislocation density at surfaces produced by ID sawing (Ref 8), and the stresses generated in wafering (Ref 9). For example, Dyer (Ref 7) has found that fracture caused by cropping, centerless grinding, exit chipping (bending), and deep saw damage results from excessive feed rates, saw blade eccentricity, and vibration. Figure 3 shows examples of a cropping chip, a crack produced by centerless grinding, a chip produced during wafering, and exit (bending) chips in a (100) silicon ingot. Data indicate that such cracks are either due to bending or Hertzian stress generation and are generally produced at the edges of the ingot or wafer. A fracture mechanics analysis (Ref 9) of the ID wafering geometry in which the ingot is rotated during the cut and the center 25 mm ($\frac{1}{4}$ in.) is pulled away has shown that the minimum wafer thickness depends on the surface damage (cracks) and the stresses imposed by the blade. The minimum wafer thickness is given by:

$$t^2 = \sqrt{\pi a_c P_A \beta / K_{Ic}} \quad \textbf{(Eq 1)}$$

where t is the water thickness, a_c is the minimum surface damage (crack), K_{Ic} is the fracture toughness, P_A is the applied force, and β is a parameter related to the ingot geometry. Equation 1 predicts that the minimum wafer thickness under this condition may be controlled to 200 μm for an ingot diameter of 100 mm (4 in.) when the minimum flaw size is less than 8 μm. The minimum flaw size is related to the subsurface damage when the diamond grit contacts the silicon.

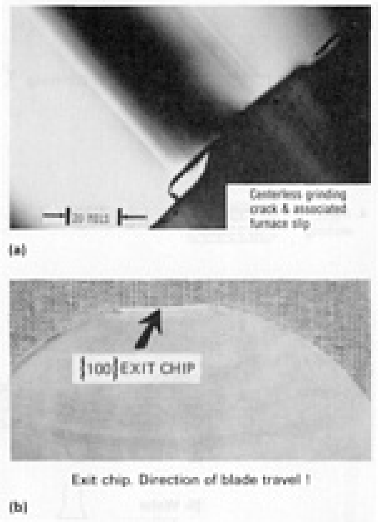


Fig. 3 Cracks induced by Hertzian stresses at the edges of an ingot or water. (a) Cracks and cavities associated with slip lines. (b) Exit chip

The interaction of the lubricant with the wafer surfaces has also been studied in relation to blade parameters (Ref 10). It was found that dielectric oil produced a greater density of dislocations compared with deionized water or methyl silane, and that these effects could be linked to blade vibration.

Mechanical Damage at Silicon Surfaces Caused by Dicing

Dicing refers to the sawing of a wafer to remove completed electronic circuits (dies) prior to packaging. The wafer is mounted on a vacuum chuck that is part of a movable stage. A rotating circular blade impregnated with diamond grit on the outside edge is brought into contact with the wafer surface while being sprayed with a lubricant. Typically, a linear cut is made along the paths separating the devices. Figure 4 shows scanning electron micrographs (SEMs) of the damage produced at the perimeter of a silicon die from a typical circuit. This die was diced by a diamond-impregnated blade using water as the lubricant. Figures 4(a) and 4(b), at higher magnification, show damage, evident as conchoidal fracture, as well as plastic smearing of the cut surface just beneath the top surface of the die.

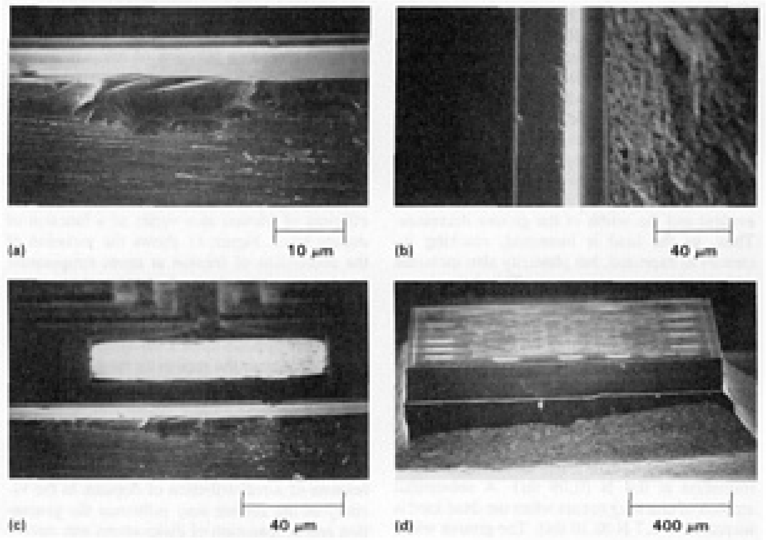


Fig. 4 Scanning electron micrographs showing typical dicing damage at the perimeter of a silicon die from an LSI circuit

Dicing damage has been studied in laboratory experiments, where blade speed, silicon temperature, and lubricant environment could be controlled (Ref 11). Figure 5 shows SEMs of a typical groove formed in a (100) silicon surface by a blade rotating at 1000 rev/min. Deionized water was used as the lubricant, and the depth of the groove and the silicon temperature were controlled. Figure 5(a) shows that the surface of the groove is streaked with even finer grooves. The higher-magnification insets (Fig. 5b to d) show that the morphology of these small grooves varies with the silicon temperature. As the temperature is increased, these miniature grooves become broader, and each can be associated with an individual diamond grid particle that contacted the surface.

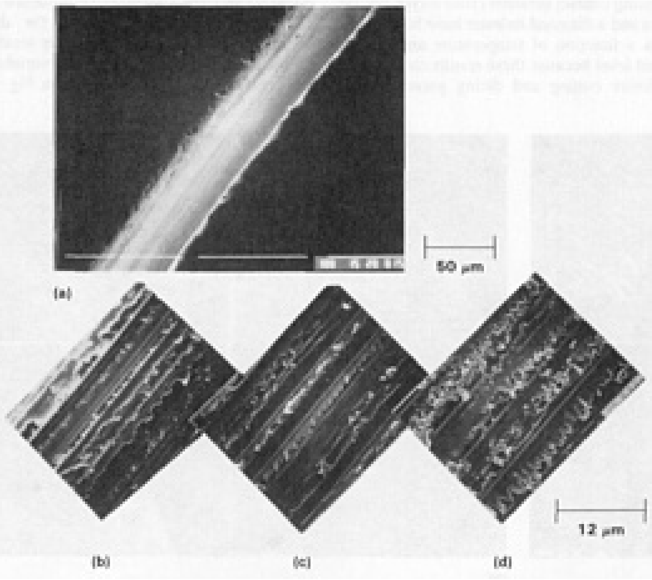


Fig. 5 Effect of temperature on groove surface morphology of (100) *p*-type silicon. (a) Typical groove produced by blade rotating at 1000 rev/min using deionized water as lubricant. (b) Room temperature. (c) 150 °C (300 °F). (d) 200 °C (390 °F)

Cracks also propagate beneath the surface. An example of such crack propagation associated with the laboratory-simulated dicing groove is presented in Fig. 6, which shows that median cracks have propagated in the [110] cleavage direction, perpendicular to the orientation of the dicing groove.

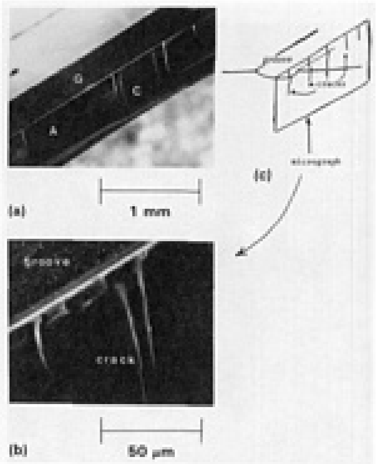


Fig. 6 Cross section of dicing groove in *p*-type silicon. (a) Scanning electron micrograph of groove (G) and subsurface cracks. (C). (b) Close-up view. (c) Schematic showing location of subsurface damage beneath groove

Simulation of Dicing Damage by Single-Point Diamond Scratching of Silicon

Laboratory simulation experiments of dicing damage in silicon have shown that the subsurface damage results from the contact of individual diamond grit particles. As a result, studies have been carried out by simulating single-point contact of silicon surfaces using diamond indenters (Ref 12, 13). One example of this type of study is shown schematically in Fig. 7. An instrumented Vickers pyramidal diamond is dead-weight loaded and then traversed across the polished silicon surface in a specific crystallographic direction. The diamond generates a groove by either single, multiple, or reciprocating motion, and the silicon can be heated or immersed in a lubricant during the test. The groove contains cracks

(Fig. 7a) within an orientation that depends on wafer crystallography and scratching direction. Expected crack geometries for gallium arsenide and silicon are indicated in (Fig. 7(b).

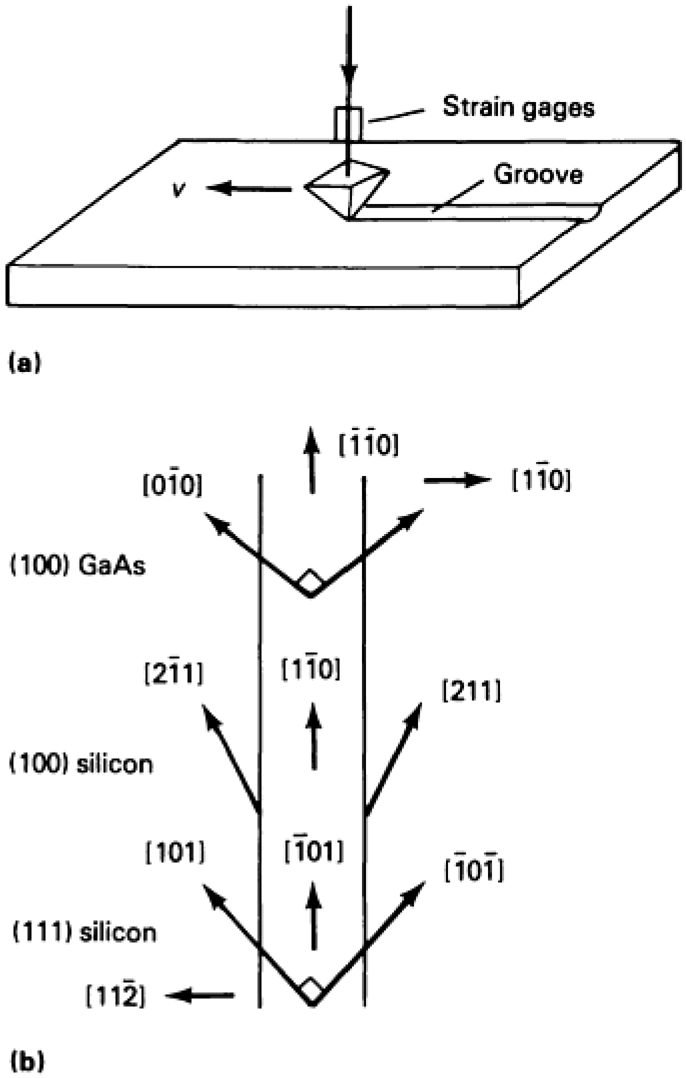


Fig. 7 Effect of crystallography on scratch testing with an instrumented Vickers pyramidal diamond indenter. (a) Schematic of test apparatus. (b) Cracking geometries generated in (100) and (111) silicon and (100) gallium arsenide

Figure 8 shows scanning electron micrographs of the scratch morphology and the subsurface damage as a function of temperature and the dead-weight load on the diamond. When the load was 0.4 N (0.09 lbf), the room-temperature scratch indicates long, lateral cracks that generate large chips near the surface. The cross-sectional view shows long, branched median cracks that propagate in the [110] direction. The extent of this damage is increased when the load is increased to 0.7 N (0.16 lbf). The surface morphology changes when the temperature is increased. For example, the scratch made at 100 °C (212 °F) shows evidence of plastic deformation in the groove region and a reduction of subsurface cracking. As the temperature is increased to 300 °C (570 °F), the plasticity is more evident and the width of the groove decreases. Thus, as the load is increased, cracking increases as expected, but plasticity also increases with increasing temperature. This behavior is indicative of a brittle-to-ductile transition.

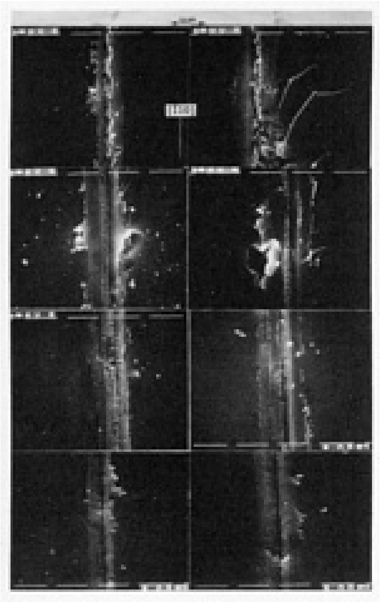


Fig. 8 Scanning electron micrographs of single scratches in (100) *n*-type silicon as a function of scratching temperature and load on the diamond. (Right side) Vickers pyramidal diamond at 0.4 N (0.09 lbf). (Left side) Vickers pyramidal diamond at 0.7 N (0.16 lbf). Under both load conditions, silicon temperature was at room temperature, 100 °C (212 °F), 200 °C (390 °F), and 300 °C (570 °F), respectively (top to bottom).

Similar experiments have been carried out on gallium arsenide surfaces. Figure 9 compares the scratch morphology at (100) *n*-type silicon and (100) *p*-type gallium arsenide surfaces scratched at 200 °C (390 °F) by a dead-weight loaded Vickers diamond using loads of 0.4 and 0.7 N (0.09 and 0.16 lbf). It is evident that the gallium arsenide surface also exhibits plastic deformation at 0.4 N (0.09 lbf). A substantial amount of cracking occurs when the dead load is increased to 0.7 N (0.16 lbf). The groove width is at least three times larger for the gallium arsenide surface compared with the silicon surface.

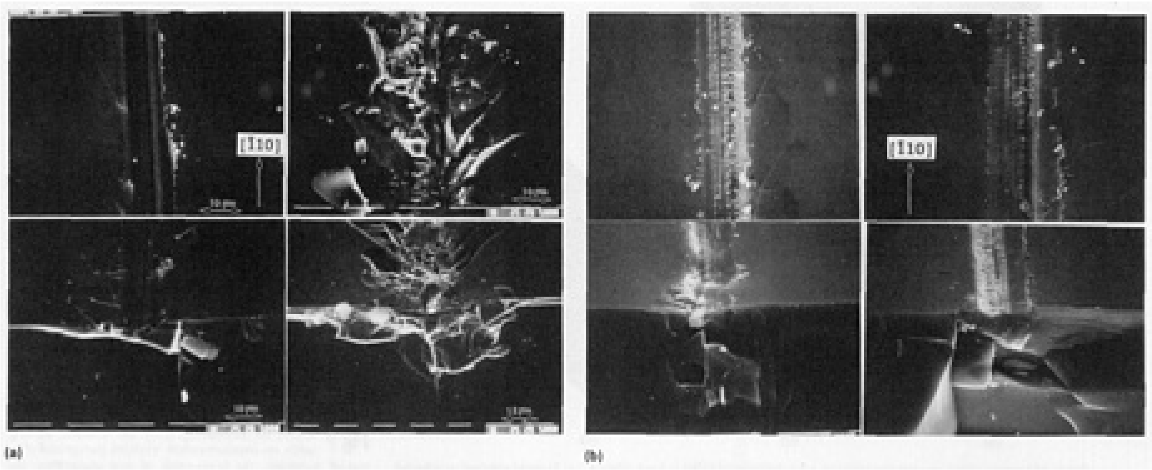


Fig. 9 Scanning electron micrographs of single scratches made by a Vickers pyramidal diamond at 200 °C (390 °F) as a function of material and load on the diamond. (a) *p*-type gallium arsenide. (b) *n*-type silicon. For both materials, load was 0.4 N (0.09 lbf) (left side) and 0.7 N (0.16 lbf) (right side).

Dynamic Friction Coefficient in Silicon and Gallium Arsenide as a Function of Temperature and Doping

The dynamic friction coefficients associated with sliding contact between (100) *n*-type silicon surfaces and a diamond indenter have been measured as a function of temperature and dopant type and level because these results can be used to optimize cutting and dicing parameters of these semiconductors. Figure 10 shows the friction coefficients obtained with a Vickers pyramid diamond as a function of temperature for (100) *n*-type silicon with a resistivity of 3.3 $\Omega \cdot$ cm. The dead-weight load on the diamond was varied from 0.4 to 0.7 N (0.09 to 0.16 lbf). The friction coefficient exhibits a maximum with temperature at approximately 200 °C (390 °F). As the load is increased, the variation in the coefficient of friction is also increased. The coefficient of friction also varies as a function of dopant level. Figure 11 shows the variation of the coefficient of friction at room temperature versus resistivity obtained at room temperature by sliding a Vickers diamond in the [110] direction; the load on the diamond was 0.5 N (0.11 lbf). There appears to be a variation of approximately 15% over the resistivity range of 10^{-2} to 10 $\Omega \cdot$ cm.

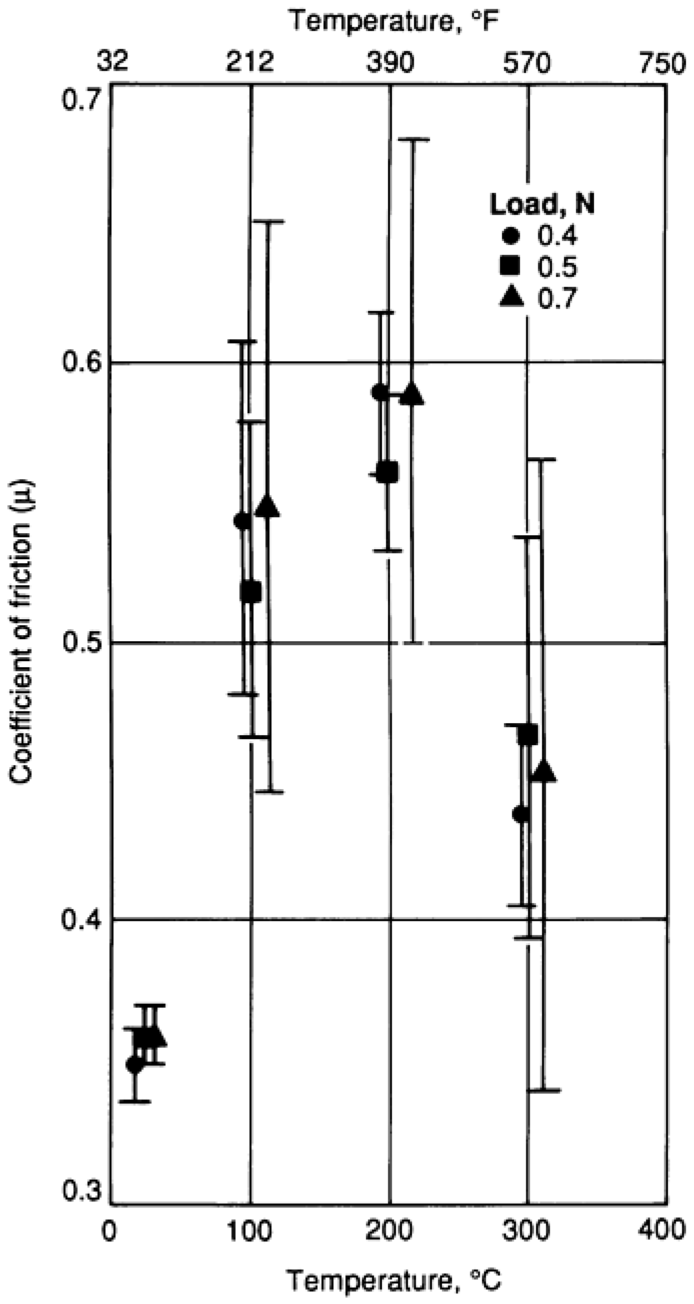


Fig. 10 Plot of the dynamic coefficient of friction versus temperature for (100) *n*-type silicon (electrical resistivity, ρ, of 3.3 $\Omega \cdot$ cm) abraded by a Vickers pyramidal diamond at selected loads.

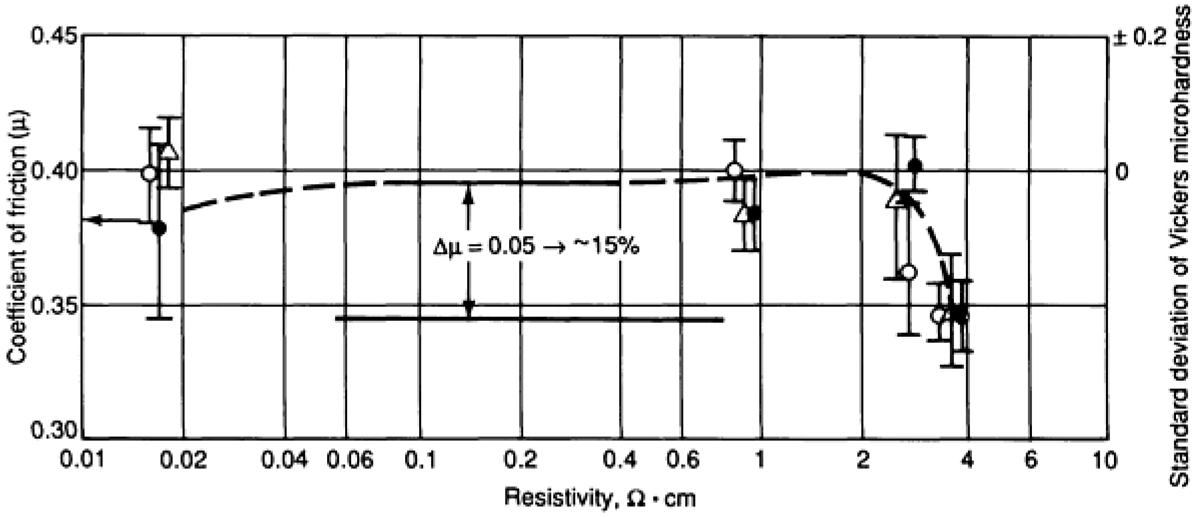


Fig. 11 Plot of the coefficient of friction versus electrical resistivity for (100) *n*-type silicon abraded at room temperature by a Vickers pyramidal diamond with a 0.5 N (0.11 lbf) load

The variability of the coefficient of friction with dopant concentration is not completely understood, but there has been some suggestion (Ref 14) that surface space charges that form because of a redistribution of dopants in the vicinity of the surface may influence the generation and propagation of dislocations and microcracks. This model can be described as follows: Semiconductor surfaces have electrical fields that physically extend into the bulk. The depth of these electrical fields is described by a Debye length that is related to dopant concentration, temperature, and other surface effects such as surface states. When the dynamic coefficient of friction is measured, compressive and tensile stresses are generated ahead and behind the sliding diamond. If the physical extent of the stresses is of the order of the space charges, then dopants can have a measurable influence on the coefficient of friction because the damage caused by sliding will be retarded or accelerated by the surface space charges. The variability of the coefficient of friction shown in Fig. 11 implies that the surface becomes harder, damage is reduced, and electrical fields are stronger as the doping is increased.

References

1. W.R. Runyan, *Silicon Semiconductor Technology,* McGraw-Hill, 1974
2. J.J. Wortman and R.A. Evans, *J. Appl. Phys.,* Vol 36 (No. 1), 1965, p 153
3. J.E. Sinclair and B.R. Lawn, *Proc. R. Soc. (London) A,* Vol 329, 1972, p 83
4. See, for example, the Proceedings of the International Symposium on Dislocations in Tetrahedrally Coordinated Semiconductors, *J. Phys. (Orsay),* Coll. C-6, Suppl. 6, 1979
5. R.H. Glaenzer and A.G. Jordon, *Solid-State Electron.,* Vol 12, 1969, p 247
6. N. Jacksen, *Solid State Technol.,* Vol 105, July 1985
7. L.D. Dyer, *Emerging Semiconductor Technology,* STP 960, D.C. Gupta and P.H. Langer, Ed., ASTM, 1986
8. T.S. Kuan, K.K. Shih, J.A. Van Vechten, and W.A. Westdorp, *J. Electrochem. Soc., Solid State Sci. Technol.,* Vol 127 (No. 6), 1980, p 1387
9. C.P. Chen and M. Leipold, *J. Electrochem. Soc., Solid State Sci., Technol.,* Vol 131 (No. 1), 1984, p 206
10. D.S. Lim and S. Danyluk, *J. Mater. Sci.,* Vol 20, 1985, p 4084
11. J.M. Kim and S. Danyluk, *J. Mater. Sci.,* Vol 22, 1987, p 1987
12. S.H. Kim and S. Danyluk, *J. Mater. Sci.,* Vol 28, 1990, p 4892
13. H.J. Leu and R.O. Scattergood, *J. Mater. Sci.,* Vol 23, 1988, p 3006
14. S. Danyluk and S.-W. Lee, *J. Appl. Phys.,* Vol 64 (No. 8), 1988, p 4075

Introduction to Materials for Friction and Wear Applications

W.A. Glaeser, Battelle Columbus Laboratories

Introduction

WEAR OF TOOLS AND WEAPONS dates back to the dawn of human history. Beginning with stone and wood, materials at hand were adapted for increasingly complex uses. The search then began for harder, more wear-resistant rocks and materials (such as flint and obsidian) that could be shaped to a sharp edge and that would hold an edge. This approach to materials selection has not changed much right up to today. Very few materials have been developed for specific tribological applications; instead, available materials have been adapted. The development of bronze and then cast iron and steel quickly led to their use in applications involving sliding and rolling contact. Perhaps the first attempt to design a wear-resistant tool involved embedding stones in wooden plows. Examples of this advance in agriculture have been found that date from the sixteenth century. This practice continued on into the nineteenth century; examples of such plows have been found in remote regions of the American Southwest, where they were used in place of iron plows because of the difficulty of obtaining processed iron.

With the industrial revolution came the need for low-friction bearings and pistons for steam engines and water-powered machines. The use of lubricants became more sophisticated, and the need for effective journal bearings demanded the use of soft bearing materials such as lead and tin. The babbitt alloys that were developed for this need might be considered *bona fide* tribological materials. Still, however, most materials used for the new machinery were adapted--as were, for example, cast iron, bronze, and wood.

Selection Guidelines

The following sections on materials for friction and wear applications provide information on a wide variety of materials that have potential tribological use. However, as mentioned earlier, very few engineering materials have been uniquely developed for tribologists. How then does one go about selecting the proper material for a given application? Following are some general guidelines that might be used to reduce time in the selection process.

For wear and abrasion resistance, the general rule is, the harder the better. Thus, for the most economical approach, one would select heat-treated carbon or alloy steels. Other hard materials, such as cemented carbides, tool steels, and hardfacing alloys, are expensive and should be reserved for severe abrasion conditions or special environmental requirements.

For conditions of marginal lubrication, the cast irons are an economical choice. Gray cast iron has been used in many applications requiring resistance to adhesive wear (for example, piston rings, cylinder liners, wheels, slideways, and thrust bearings). Cast iron also provides good damping properties for the control of machine vibrations.

For prevention of galling and seizure, microstructure is as important as hardness. Materials containing large carbides, such as hardfacing alloys and alloy cast irons, can be considered for applications where this is a factor.

Bearing Applications. Rolling-contact bearings use special-grade steels designated as "bearing steels." For high-temperature applications, stainless steels, tool steels, cemented carbides, superalloys, and ceramics are useful in sliding or rolling contacts. If corrosion resistance is required, stainless steels, hardfacing alloys, superalloys, titanium-base alloys, ceramics, carbon-graphite, and polymers should be considered.

Journal bearing materials, which require conformability, embeddibility, softening under frictional heating, and fail-safe properties, are discussed in the article "Friction and Wear of Sliding Bearing Materials." These materials include babbitts, bearing bronzes, zinc-base alloys, and aluminum-base alloys; some polymers also fall under this classification. For sliding bearings operating in water, carbon-graphite is an excellent selection; some polymers are appropriate for these

applications as well. Self-lubricating materials for use under dry sliding conditions include various polymers, carbon-graphite materials, and metal-matrix composites.

Material State

Hardness often is a necessary property for wear resistance. Similarly, material physical properties such as thermal conductivity, thermal expansion, density, and heat capacity are important parameters in the selection of materials. However, the state in which one finds materials when they are to be assembled into a machine is equally important. This means the microstructure, the surface finish, the residual stresses, and the porosity and depth of hard coating or soft coating can be significant factors in the final performance of a bearing, gear, or seal. Thus, just changing from one material to another will not necessarily solve an excessive wear or surface damage problem.

For example, a roller bearing with races heat treated and finish ground with a condition of residual tensile stresses in the race surfaces will have a propensity to spalling failure. Thus, when this bearing experiences contact stresses above its normal operating limits (as sometimes happens in machinery), it is likely to fail. A bearing with compressive residual stresses in the races has added insurance against failure when unplanned excursions of load occur.

Cast irons that have the same hardness but different graphite structures will behave differently in wear. The preferred graphite structure in cast irons is a well-distributed flakelike morphology. The size and distribution of flakes is a significant factor in the wear resistance of the cast iron.

Steels for rolling-contact bearings must be exceptionally clean (that is, free of inclusions). Hard silicates are the most damaging to performance. Special vacuum melting processes are used to produce high-quality bearing steels. Thus, in this class of materials, hardness alone is not the deciding parameter for optimum performance.

Other material properties that have nothing to do with the particular tribological problem should also be considered. For example, a shaft may have to withstand bending fatigue conditions or stress-corrosion as well as wear of the contacting bearing material. It may turn out that tensile strength is as important as hardness for a given application.

Tribological problems such as excessive wear are not always solved by searching for a better material. Sometimes appropriate design may be the more economical and lasting solution. Designs that provide a lubricant film support to a contacting system may mitigate the need for a super wear-resistant material.

Surface finish is also of great importance in tribological applications. Smooth surfaces are generally required. This is especially important in journal bearings and rolling-contact bearings, where very thin lubricant films are needed for the performance of the bearings. If the surfaces are too rough, the high points penetrate the thin oil film and initiate wear or failure.

Great savings in manufacturing cost can be achieved by applying coatings or surface treatments to economical structural steels in order to provide wear resistance at selected areas in machine components. The Section "Surface Treatments and Coatings for Friction and Wear Control" in this Volume discusses the selection of wear-resistant coatings and surface treatments.

The following articles on materials for friction and wear applications have been written by experts in each class of materials. Because of the individual style and experience of each author, you will not find a constant format running through each article. This is to be expected, however, because each class of material has a different history and significance in its application to tribology.

Friction and Wear of Cast Irons

T.S. Eyre, Eyre Associates (U.K.)

Introduction

CAST IRON is a general term applied to a family of engineering materials that includes gray iron, white or chilled iron, malleable iron, and nodular iron. These irons have very different compositions, casting characteristics, and heat treatments, which result in physical, chemical, and mechanical properties of wide variation.

Cast irons are unique tribological materials that can be used in a range of applications, particularly those in which wear resistance is important. The matrix of these materials can be varied from pearlite to martensite, in order to support graphite, which is a solid lubricant. Hard phases can be varied to resist abrasion. The primary characteristics of gray and nodular irons are described below, in terms of very general selection guidelines, as well as later in this article, in greater detail.

Gray Iron. All of the excess carbon in gray iron is present in the form of flakes of graphite in a matrix that varies from all-ferrite to pearlite, with or without phosphide eutectic or carbides. Gray irons are brittle and have relatively low strength and hardness properties. Graphite provides good damping capacity, and its thermal conductivity value is higher than that for steel. Graphite cavities provide chip-breaking qualities, which allow free machining with short chips. A fractured surface will appear gray and dull, because of the presence of graphite.

Because graphite is an excellent solid lubricant, it confers good sliding wear resistance with a relatively low coefficient of friction. This class of cast iron is widely used for internal combustion engine components, including engine blocks, cylinder liners, and piston rings, as well as machine tool slideways.

White or Chilled Iron. Carbon can be retained as a primary carbide in a pearlitic or martensitic network by suitable alloying or heat treatment, or by the use of chilling. In this condition, the iron is very hard (up to 700 HV), unmachinable, and quite brittle. A fractured surface will appear white and crystalline. By varying the manufacturing route (for example, by chilling), it is possible to produce components that are gray, or graphitic, in some areas and white, or carbidic, in the chilled areas. These materials are used for abrasive wear applications in mining, grinding, mineral handling equipment, and camshafts in the internal combustion engine.

Malleable Iron. Castings initially produced as white iron can be suitably heat treated to produce a more ductile structure. Upon fracture, this structure produces either a white facet (whiteheart), which is due to the presence of pearlite, or a black facet (blackheart), which is due to the presence of graphite. These treatments impart a significant change to the otherwise brittle characteristic of a white iron. Ductility and impact properties improve considerably for stress-bearing components for which gray iron would be unsuitable.

Nodular or Spheroidal Graphite Iron. By converting graphite from the flake to a nodular form, the brittleness of flake (gray) irons is immediately diminished and, with suitable heat treatment, a steel-like structure can be produced. Because pure materials and rather special melting procedures are required, cost is increased significantly, but the improvement in mechanical properties allows these irons to compete with heat-treated steels for applications such as crankshafts and gears. Significant developments in recent years have extended the properties of specific alloys, such as austempered ductile iron, which competes favorably with surface-hardened steel in applications such as cast gears.

Constitution of Cast Iron

Cast iron is essentially an iron-carbon alloy that contains from 2.5 to 4.0% carbon and other essential alloying elements, the most important of which are silicon and phosphorus. Pure iron melts at 1535 °C (2795 °F); the eutectic of iron and carbon occurs with a carbon content between 4.2 and 4.3% and a melting point of 1153 °C (2107 °F). The eutectic melting temperature for cast iron varies slightly with silicon content, and usually falls between 1135 and 1150 °C (2075 and 2102 °F). Cast irons therefore have a melting range that is approximately 400 °C (720 °F) lower than that for steel.

Additions of silicon lower the percentage of carbon in the eutectic by about 0.33% for each 1% of silicon added, so that with 1% of silicon, the eutectic will occur at approximately 3.97% of carbon. Additions of phosphorus affect the eutectic carbon value to about the same amount as silicon, so that for an iron containing 1% silicon and 1% phosphorus, the eutectic will occur with approximately 3.6% carbon. Other elements, in the amounts that are normally present, have a much smaller effect on eutectic carbon value, but they do have other significant effects that are referred to later in this article.

A cast iron can have a composition that makes it hypoeutectic, eutectic, or hypereutectic with regard to carbon. Its position in this respect is important, because not only does the initial solidification point (liquidus) depend on it, but also the manner of solidification and, thereby, the structure.

The carbon content of most commercial irons is within the range noted above (from 2.5 to 4.0%). The iron-carbon diagram in Fig. 1 shows only the iron carbide/austenite (ledeburite) system in the high-carbon region. Such irons, when commercially produced, are known as white, or chilled, irons and are hard, brittle, and unmachinable. Gray irons, however, solidify with the formation of graphite instead of iron carbide, and are reasonably strong and machinable. An intermediate material, known as mottled iron, in which both graphite and iron carbide are formed during solidification, also exists. Such irons are harder and less readily machinable than the gray irons.

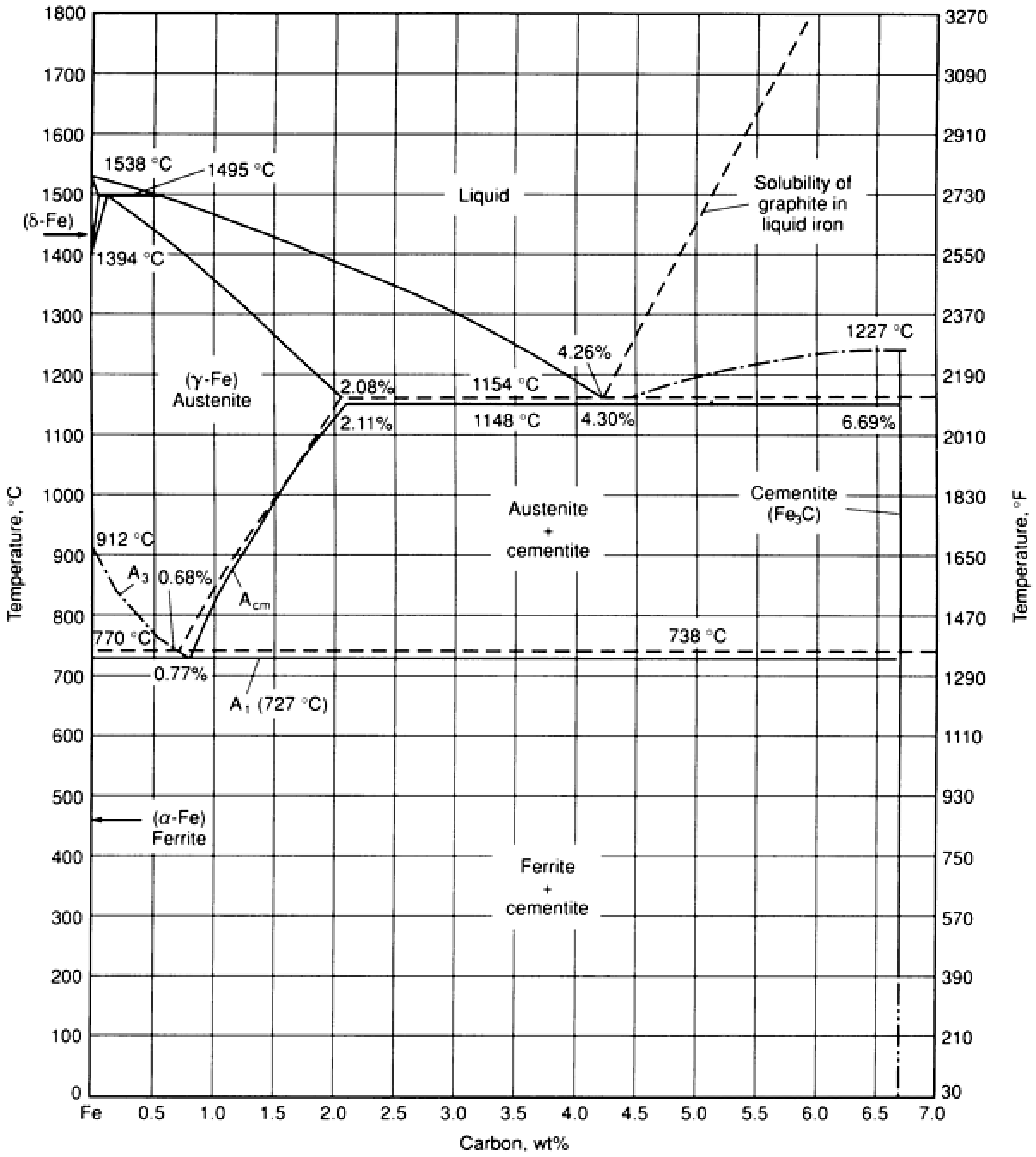


Fig. 1 Iron-carbon diagram, where solid curves represent metastable system Fe-Fe_3C, and dashed curves, the stable system Fe-graphite

Whether an iron solidifies as gray, white, or mottled primarily depends on the eutectic value and the rate of cooling. However, before attempting to evaluate the effect of cooling rate, the effect of composition must be simplified. The two most important factors in understanding the mechanical properties and, thus, their efficiency in engineering applications, are:

- Composition, or carbon equivalent value
- Cooling rate, or section sensitivity

Both factors act independently and have a significant effect on the microstructure, which itself controls the properties produced. Different microstructures can be produced at different positions in the same casting.

Composition, or carbon equivalent (CE) value, is the most common simplification used, in which

$$CE = \text{Total carbon \%} + \frac{\text{Silicon\%} + \text{Phosphorus\%}}{3} \quad \textbf{(Eq 1)}$$

It is this value, rather than the actual total carbon content, that applies to commercial irons when the iron-carbon diagram in Fig. 1 is used.

The CE value immediately indicates whether an iron is hypoeutectic or hypereutectic, and by how much. Thus, an iron with 3.2% carbon, 2% silicon, and 0.4% phosphorus has a CE value of 4.0 and is hypoeutectic. An iron with 3.2% carbon, 2% silicon, and 1.3% phosphorus has a CE value of 4.3 and is eutectic. An iron with 3.2% carbon, 2.9% silicon, and 1.3% phosphorus has a CE value of 4.6 and is hypereutectic.

In general, the lower the CE value, the greater is the tendency for an iron to solidify as either white or mottled. The solidification period determines whether a given iron becomes gray, mottled, or white; the form of the graphite is altered very little by the subsequent cooling period.

Cooling Rate, or Section Sensitivity. The tensile test can be regarded as the standard test by which cast iron is specified, the exception being white iron, for which hardness only is specified. Tensile strength is the stress required to pull apart a testpiece by an axially applied load. The test is almost always performed on a round, machined testpiece, in which the middle section of the length is reduced in diameter. The testpiece is machined from either independently cast bars or broken transverse test bars. Specifications normally give details of casting sizes appropriate to each size of cast bar. Because the strength of cast iron is particularly dependent on the section size into which it is cast, the strength of gray iron is defined in relation to the diameter of as-cast bars from which standard tensile specimens can be machined.

For gray flake iron, each of the eight grades is designated by the tensile strength required on the 30 mm (1.2 in.) diameter bar (Table 1). The specification for each grade, however, requires a higher tensile strength for bars of smaller diameter, and a lower tensile strength for bars of larger diameter. However, the variation in strength given in the specification according to size of the test bar is less than has been found in practice, and merely indicates in an arbitrary manner the commercial difficulty of producing a given tensile strength in a given bar.

Table 1 Expected tensile strengths of testpieces machined from cast-on test samples and from the castings

Grade	Casting, over:		Section thickness, up to and, including:		Tensile strength			
					Cast-on test samples		**Castings**	
	mm	**in.**	**mm**	**in.**	**MPa**	**ksi**	**MPa**	**ksi**
100	2.5	0.1	10	0.4	. . .	. . .	120	17.4
	10	0.4	20	0.8	. . .	. . .	90	13.1
150	2.5	0.1	10	0.4	. . .	. . .	155	22.5
	10	0.4	20	0.8	. . .	. . .	130	18.9
	20	0.8	40	1.6	120	17.5	110	16.0
	40	1.6	80	3.2	110	16.0	95	13.8
	80	3.2	150	6.0	100	14.5	80	11.6
	150	6.0	300	12.0	90	13.0	. . .	. . .
180	2.5	0.1	10	0.4	. . .	. . .	185	26.8
	10	0.4	20	0.8	. . .	. . .	160	23.2
	20	0.8	40	1.6	150	21.8	135	19.6
	40	1.6	80	3.2	135	19.6	115	16.7
	80	3.2	150	6.0	125	18.1	100	14.5
	150	6.0	300	12.0	100	14.5	. . .	. . .
200	2.5	0.1	10	0.4	. . .	. . .	205	29.7
	10	0.4	20	0.8	. . .	. . .	180	26.1
	20	0.8	40	1.6	170	24.7	155	22.5
	40	1.6	80	3.2	150	22.8	130	18.9
	80	3.2	150	6.0	140	20.3	115	16.7

	150	6.0	300	12.0	130	18.9	. . .	. . .
220	2.5	0.1	10	0.4	. . .	. . .	220	31.9
	10	0.4	20	0.8	. . .	. . .	195	28.3
	20	0.8	40	1.6	175	25.4	170	24.7
	40	1.6	80	3.2	165	23.9	145	21.0
	80	3.2	150	6.0	150	21.8	130	18.9
	150	6.0	300	12.0	140	20.3	. . .	. . .
250	5	0.2	10	0.4	. . .	. . .	250	36.3
	10	0.4	20	0.8	. . .	. . .	225	32.6
	20	0.8	40	1.6	210	30.5	195	28.3
	40	1.6	80	3.2	190	27.6	170	24.7
	80	3.2	150	6.0	170	24.7	155	22.5
	150	6.0	300	12.0	160	23.2	. . .	. . .
300	10	0.4	20	0.8	. . .	. . .	270	39.2
	20	0.8	40	1.6	250	36.3	240	34.8
	40	1.6	80	3.2	220	31.9	210	30.5
	80	3.2	150	6.0	210	30.5	195	28.3
	150	6.0	300	12.0	190	27.6	. . .	. . .
350	10	0.4	20	0.8	. . .	. . .	315	45.7
	20	0.8	40	1.6	290	42.1	280	40.6
	40	1.6	80	3.2	260	37.7	250	36.3
	80	3.2	150	6.0	230	33.4	225	32.6
	150	6.0	300	12.0	210	30.5	. . .	. . .

Source: ISO R185 · 1961; British Standard BS 1452:1990

Tensile Properties

All of the principal countries specify cast iron on the basis of standard-diameter test bars. However, it is common knowledge in the foundry industry that test bar results cannot be used to give the strength of the casting itself, because section thicknesses and cooling rates can vary appreciably. This can be summarized as follows:

- For all foundry alloys, the separately cast test bar is essentially a check on metal quality only
- Strength relationships between the test bar and the casting vary with the nature of the alloy being cast
- The control of castings, especially for vital services, demands the use of the separately cast test bar as a check on metal quality, but must be supported with check tests on the castings themselves. For repetition castings, such tests can be conducted on a proportion of the castings produced. Otherwise, some suitable form of nondestructive testing must be applied

It is important to realize that national specifications merely define the metal in the ladle and do not give a direct indication of the strength of the casting itself. As explained earlier, the strength of the material in the casting is a function of the cooling rate of the particular portion under consideration. However, when the strength of the metal cast under standard conditions is known, it becomes possible to estimate, with some degree of accuracy, the probable tensile strength to be found in a gray iron casting of a given thickness. If the test bar were cast onto the casting itself, then it would be impossible to relate the results obtained from such a test bar with those obtained from another test bar under standard conditions. Such a test would be almost valueless, because the test bar would not be representative of the casting itself, nor could it be related to other standards.

In past practice, it was quite common to specify cast-on test bars, on the assumption that such a test bar must necessarily be representative of the casting itself. However, this practice did not provide the necessary information. Except in certain cases, the use of the independently cast test bar has now almost entirely superseded the use of the cast-on test bar.

Effect of Section on Strength. To interpret the results obtained from the standard test bar in terms of the strength of the casting itself, some relationship must be established between various section sizes and the bar. This is best exemplified by discussing the properties of gray iron.

A statistical survey of a large number of industrially cast gray iron test bars of different sizes has been made, the results of which are shown in Fig. 2. The curve shows the falling off in strength with increasing test bar size for the main grades of unalloyed gray cast iron. The line for each grade is really the center of a zone of uncertainty amounting to 5 to 10 Pa (0.00075 to 0.0015 psi), because there is a possible variation in each grade that is due to differences in composition, pouring technique, and other factors. A version of this curve is usually included in the relevant national or international specifications.

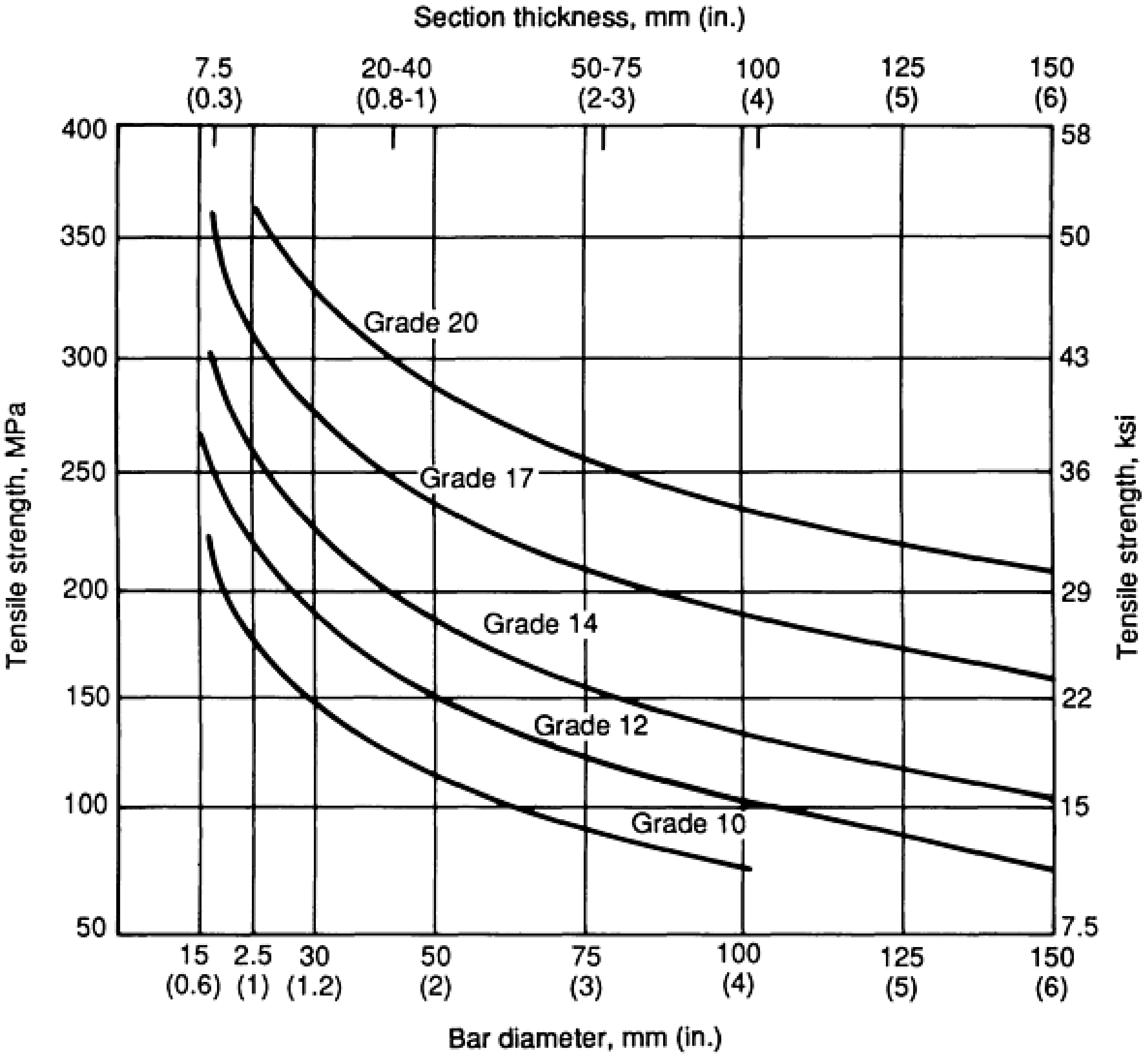


Fig. 2 Relationship between tensile strength and section thickness for different grades of flake gray iron. Source: ISO R185. 1961; British Standard (BS) 1452: 1990

The total strain (recoverable and nonrecoverable) of gray iron tensile testpieces at fracture usually varies between 0.5% for high-tensile iron and 0.75% for low-tensile iron. The non-elastic strain varies between 0.2% for high-tensile iron and 0.6% for low-tensile iron.

White Irons

If the iron is sufficiently below the eutectic value, has a very low silicon content, and contains appreciable quantities of carbide or stabilizing elements, or if the cooling rate is sufficiently rapid, then the formation of graphite does not occur. Under these conditions, solidification takes place first by the formation of austenite dendrites. The remaining liquid in the interdendritic spaces then becomes enriched in carbon and solidifies as a eutectic of iron carbide and austenite, known as ledeburite. As the solid casting cools down, at a temperature ranging from 720 to 750 °C (1330 to 1380 °F), the austenite, which is γ iron with carbon in solution, transforms to α iron, in which carbon is substantially insoluble. Carbon is then

rejected as iron carbide in the form of lamellae alternating with lamellae of α-iron to form a matrix of pearlite. The initially formed iron carbide, however, remains unchanged.

The structure of a white iron at room temperature, therefore, consists of primary dendrites of pearlite with interdendritic areas of transformed ledeburite, which is a eutectic structure of iron carbide and pearlite. The pearlite areas are formed from the original austenite. In some cases, the austenite produced during the final eutectic solidification of iron carbide and austenite is deposited on the already existing austenite dendrites, and the interdendritic area is predominantly carbide. The final structure will then consist of dendrites of pearlite surrounded by cementite that contains very little transformed eutectic pearlite (Fig. 3).

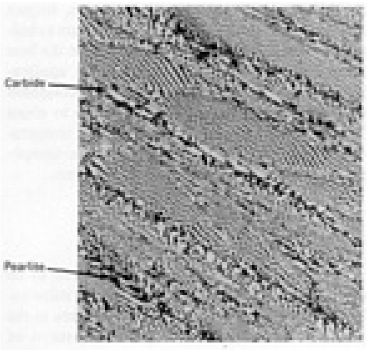


Fig. 3 Pearlitic white iron showing the original austenite dendrites that transformed to pearlite and the interdendritic eutectic structure of cementite and austenite, where the austenite transformed to pearlite. White phase, carbide; dark phase, pearlite. 80×

These irons are characterized by a completely white fracture and a high hardness (400 to 600 HV), which is an average of the carbide (900 to 1200 HV) and pearlite (200 to 300 HV). Various grades of white iron are available, depending on composition and heat treatment. The cheapest white irons are either unalloyed or low-alloy grades containing between 2.4 and 3.4% carbon and up to 2.0% chromium. A hardness of 350 HV is achieved in the cast condition. By increasing the chromium content (up to 10%), it is also possible to increase the carbon content to produce increased amounts of carbides. By including nickel, it is also possible to carry out heat treatment to produce a martensitic, rather than pearlitic, matrix.

These changes have the overall effect of increasing the hardness to 600 HV and reducing the effect of section sensitivity. When increased corrosion resistance and maximal abrasion resistance are required, irons containing up to 28% chromium can be used, but as the percentage of carbides (or carbon content) increases, impact resistance decreases. Thus, it is important to balance the composition carefully to obtain the best combination of properties for specific applications. These irons in the heat-treated condition develop the maximum hardness of up to about 700 HV, and are the least affected by temperature in service. Table 2 summarizes the composition and hardness of various white irons.

Table 2 Composition and hardness values for white irons

Composition, %									Hardness, HV
C	**Si**	**Mn**	**Cr**	**P**	**Mo**	**Ni**	**Cu**	**S**	
2.40-3.40	0.50-1.50	0.20-0.80	2.0 max	0.15 max	. . .	. . .	. . .	. . .	400
2.70-3.20	0.30-0.80	0.20-0.80	1.50-3.50	0.15 max	0.50 max	3.0-5.50	. . .	0.15 max	500
2.80-3.20	0.30-0.80	0.20-0.80	8.0-10.0	0.15 max	0.50 max	4.0-6.0	. . .	0.15 max	550
3.0-3.60	1.0 max	0.50-1.50	14.0-17.0	0.10 max	3.0 max	2.0 max	. . .	0.10 max	710
2.0-2.80	1.0 max	0.50-1.50	22.0-28.0	0.10 max	1.50 max	2.0 max	. . .	0.10 max	655

Gray Irons

In hypoeutectic irons, the first phase to be deposited from the melt is austenite, a solid solution of carbon in γ iron that crystallizes in the form of dendrites, the length and pattern of which depend on the temperature gradients. The austenite dendrites continue to grow, and the remaining liquid increases in carbon content until the eutectic temperature and concentration are attained. Solidification begins from a number of centers, each with an approximately spherical crystallization front, and simultaneous deposition of graphite and austenite occurs. Ultimately, these eutectic cells meet and consume the remaining liquid. The austenite of the dendrites and that of the eutectic becomes continuous, and the structure appears as a dispersion of graphite flakes in a matrix of austenite.

After solidification, the eutectic cell structure and the random orientation of the primary austenite dendrites throughout the structure are essentially complete. They cannot be subsequently modified either by cooling or subsequent treatment. Only in strongly hypoeutectic iron is the austenite dendrite pattern clearly visible.

Hypereutectic irons solidify by the direct formation of graphite from the metal in the form of coarse flakes that are called kish. Because of its low relative density, kish tends to rise to the surface of the melt. However, when entrapped by the metal, kish generally appears in the microstructure as characteristic long straight flakes, or, in rapidly cooled sections, as lumpy flakes. This phase is deposited over a range of temperatures, starting at the liquidus surface, until the carbon content of the remaining melt is at the eutectic concentration when a simultaneous crystallization of graphite and austenite occurs. Generally, this eutectic graphite in hypereutectic irons occurs in flake form on a finer scale than does the primary kish.

The austenite, upon cooling through the critical temperature range (from about 720 to 830 °C, or 1330 to 1525 °F), transforms to pearlite. The products of this transformation depend on the rate of cooling and composition of the material. Under normal conditions, the products will consist of pearlite, or graphite and ferrite, or mixtures of all three. Where the eutectoid transformation produces graphite and ferrite, the graphite is deposited on the already existing eutectic graphite flakes. The formation of ferrite is most likely to occur with slow rates of cooling and high silicon contents that reduce the stability of iron carbide, or with high CE values, or with fine undercooled graphite. The formation of a fully pearlitic structure is more likely to occur with moderately rapid cooling rates or low CE values.

If the iron is very rapidly cooled through the critical temperature, for example, by oil quenching, then the transformation of austenite to pearlite or ferrite can be suppressed, and a hard matrix of martensite formed. This treatment can be given deliberately when increased hardness is desired.

Graphite

The form and distribution in which graphite is deposited depends on numerous factors, such as melting temperature, nucleation, cooling speed, and others. The basic forms are flake graphite, aggregate or temper carbon, and nodules or spheroids.

A convenient standard for graphite (ASTM A247-47 and 67) defines shape, size, and distribution, and has been extended by the International Organization for Standardization (ISO Recommendation 945) to cover additional forms. The basic nomenclature and grading for flake graphite follows the generally accepted ASTM pattern.

Flake graphite can form as one of four types:

- Kish, which occurs in hypereutectic irons and can be present as straight and undistorted flakes (ASTM Type C; ISO Form I, distribution C) or very thick lumpy flakes
- Normal flakes (ASTM Type A; ISO Form I, distribution A)
- Rosette (ASTM Type B; ISO Form I, distribution B)
- Undercooled (ASTM Type D; ISO Form I, distribution D)

The undercooled, or type D graphite, is normally associated with rapid rates of cooling and is particularly common in thin sections, but can be produced by other means (such as addition of titanium). The rosette pattern is typical of slightly less rapid cooling and is not uncommon in the more rapidly cooled surfaces of castings with sections of 9.6 mm ($\frac{3}{8}$ in.) or more. Both undercooled and rosette graphite are frequently associated with ferrite, rather than pearlite, because the larger available surface area of graphite reduces the distance required for carbon diffusion during transformation from γ to α

iron in the critical temperature range. When ferrite is associated with random or coarse flake graphite, the iron is usually hypereutectic or the rate of cooling in the mold has been very slow.

Widmanstätten and mesh-form graphite can occur when lead or tellurium are present. These structures profoundly weaken the iron. When gray iron passes through the γ to α critical transformation range, graphite may be rejected from the γ phase onto existing graphite flakes and form spiky irregularities on the graphite surface.

When minimum friction and maximum wear resistance are required under dry, marginally lubricated or fully lubricated conditions, it is very important to precisely specify the microstructural requirements. These properties can be met by a requirement for Type A graphite in the size range from 3 to 5, with a pearlitic matrix, free from primary ferrite (Fig. 4). Phosphorus is desirable at a level of up to approximately 0.75%, which produces a hard, dispersed interdendritic phase that improves abrasion resistance at the expense of machinability.

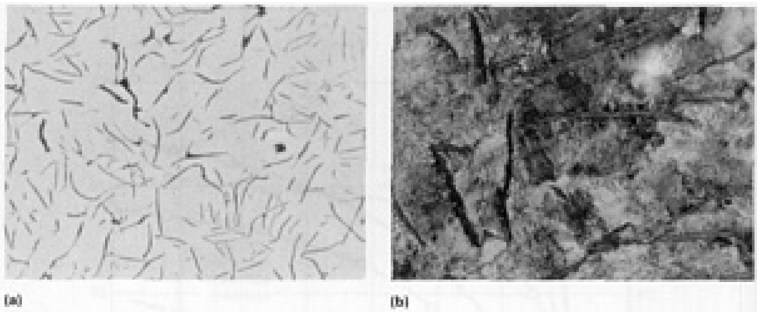


Fig. 4 (a) Gray flake iron with random graphite in an unetched matrix 75×. (b) Etching showing pearlitic matrix. 300×

Although each flake appears to be separated from adjacent ones in a two-dimensional microstructural view, it must be recognized that in each colony, as viewed in three dimensions, the flakes are connected at the nucleation source to produce a lettuce leaf-like structure. This can only be observed by deep etching the matrix and by selective dissolution, so that the graphite structure can be seen to stand above the matrix. The sharpness of the graphite cavities acts as an internal stress raiser and produces the brittleness that characterizes gray iron. However, the material is immune from the effects of external stress raisers.

Nodular Graphite. The shape of the graphite precipitated on solidification can be modified to the nodular or spheroidal form (ISO Form VI) by additions of magnesium and/or cerium. Intermediate or degenerate forms can also be produced. The ISO Form II crab type is a degenerate form of flake graphite occasionally found in magnesium-treated irons containing lead, bismuth, arsenic, or titanium, and can be suppressed by additions of cerium. Unless pure melt charge materials can be guaranteed, it is the recommended practice to use a combined addition of magnesium and cerium. Quasiflake graphite (ISO Form III) is usually the result of insufficient magnesium additions to produce the fully spheroidal form. Irregular open, or "exploded," type spheroids (ISO Form V) may occur with very pure raw materials treated with magnesium and cerium. Misshapen spheroids tend to occur in heavy sections.

Irons that contain graphite nodules (Fig. 5) that are either primarily or totally spheroidal in the as-cast condition have properties that are fundamentally different from those of gray iron with flake graphite. They possess greater tensile strength and have a considerable degree of ductility, as well as greater impact resistance and enhanced fatigue resistance. Their properties are much closer to those of steel than to gray iron (Fig. 6).

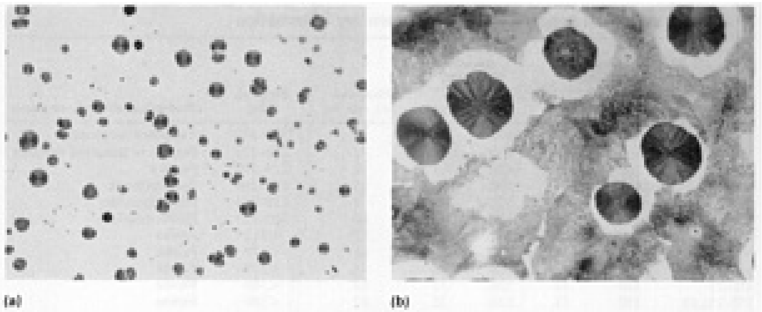


Fig. 5 (a) Nodular graphite in unetched matrix. 75×. (b) Etching shows that matrix consists of ferritic envelope around the nodules surrounded by a pearlitic matrix. 300×

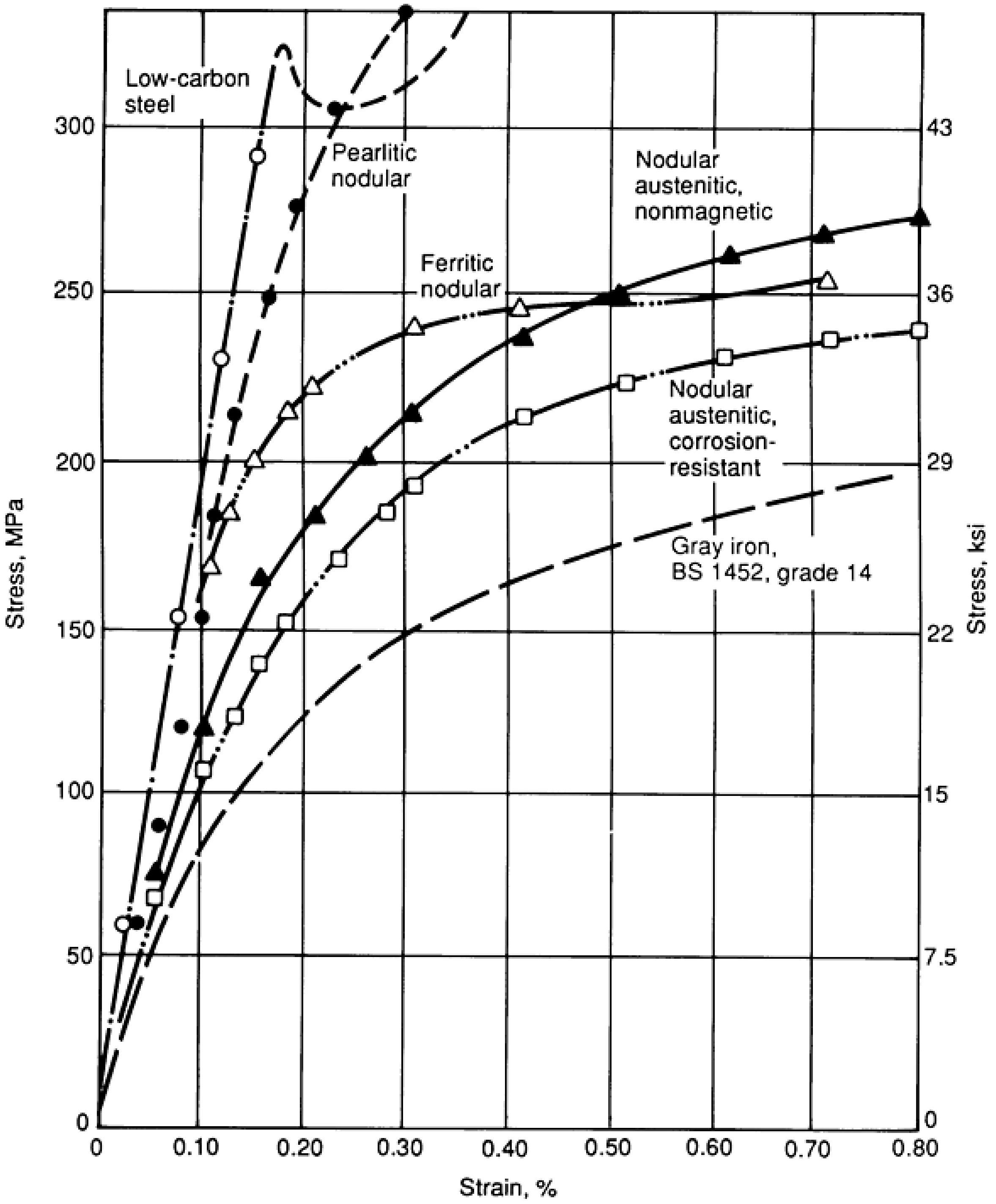


Fig. 6 Typical stress-strain curves in tension for various cast irons and steels. Source: British Standard (BS) 2789:1985

The properties to be expected in the nine different grades of nodular iron currently available are summarized in Table 3. It is also possible to heat treat these materials using techniques similar to those used for steel. In this way, a nodular graphite in a martensitic matrix can be produced.

Table 3 Mechanical properties and supplementary information

Grade	Tensile strength (R_m), min		0.2% proof stress[a]($R_{p0.2}$), min		Elongation (A), min, %	Hardness, HB[b]	Predominant structure constituent
	MPa	**ksi**	**MPa**	**ksi**			
900/2	900	130	600	87	2	302-359	Tempered martensite
800/2	800	116	480	70	2	248-352	Pearlite or tempered structure
700/2	700	102	420	61	2	229-302	Pearlite
500/3	600	87	370	54	3	192-269	Pearlite/ferrite
500/7	500	73	320	46	7	170-241	Ferrite/pearlite
450/10	450	65	320	46	10	160-221	Ferrite/pearlite
420/12	420	61	270	39	12	<212	Ferrite
400/18	400	58	250	36	18	<179	Ferrite
400/18L20	400	58	250	36	18	<179	Ferrite
350/22	350	51	220	32	22	<160	Ferrite
350/22L40	350	51	220	31	22	<160	Ferrite

Source: ISO 1083 · 1987; British Standard BS 2789:1985

(a) Verification of the 0.2% proof stress ($R_{p0.2}$), is optional.

(b) Verification of hardness values is optional.

In the as-cast state, the microstructure of nodular iron can range from a completely white structure where thin sections are involved, to that of nodular graphite in a completely ferritic matrix. By virtue of the fact that nodular irons have their own "carbon bank" available, like gray irons do, heat treatments can be applied to produce a wide range of matrix structures. Thus, by heating to a temperature above the critical temperature, graphitic carbon will dissolve in the austenite until equilibrium conditions are attained. By air-cooling from this temperature, the austenite undergoes the pearlitic transformation, thus maintaining an amount of carbon dispersed in the matrix.

On the other hand, a very slow cooling through the temperature range will deplete the carbon from the matrix, and, instead, secondary graphite is deposited around the graphite nodules. The heat treatments normally applied on a commercial scale can be summarized as follows:

- Heat to a temperature between 850 and 900 °C (1560 and 1650 °F). The length of soaking time depends on the amount of free carbide present. Air-cooling from this temperature provides a predominantly pearlitic matrix, except in heavy sections
- Remove carbides and austenitize as described above. Oil quench at a temperature between 850 and 900 °C (1560 and 1650 °F) to produce a martensitic matrix that can then be tempered to produce the required range of mechanical properties. This treatment is limited to castings of relatively thin section, unless alloying elements to improve hardenability are present
- Remove carbides and austenitize as described in the first bullet. Cool to a temperature between 680 and 700 °C (1255 and 1290 °F). It is common to hold at this temperature for a period of time in order to effect complete graphitization with the production of a ferritic matrix. The need for holding and the time involved will vary with basic composition and section size, but is usually of the order of from 8 to 12 h. The material is finally cooled to about 600 °C (1255 °F) before removal from the furnace
- The air-cooled material from the process described in the first bullet, which possesses a pearlitic matrix, can be annealed at a subcritical temperature (700 to 720 °C, or 1290 to 1330 °F), giving a completely ferritic matrix, as in the third bullet. The holding time at the subcritical temperature depends on composition and section size, but from 8 to 12 h should be adequate
- Any castings that are complex and likely to contain residual stresses, either in the as-cast condition or after normalizing, can be stress-relieved by heating to a temperature from 450 to 550 °C (840 to 1020 °F). Although no marked structural change occurs during this treatment, there is a tendency for softening to occur when the higher temperatures are used
- Nodular irons can also be transformed to a bainitic structure through the use of austempering carried out by cooling from 850 to 950 °C (1560 to 1740 °F) into a bath at a temperature between 230 and 425 °C (445 and 795 °F). Heat-treatment stresses and distortion are minimized and their overall properties

vastly improved. For example, a tensile strength ranging from 1200 to 1600 MPa (175 to 230 ksi) can be achieved with significant ductility, toughness, and fatigue strength. In this condition, austempered ductile iron is now being used for gears and other highly stressed parts

Nodular irons are used in such applications as crankshafts and gears, where they compete against through-hardened and surface-hardened steels. Their near-net-shape forming characteristics, coupled with the presence of graphite (which confers sliding wear resistance), offer a unique combination of properties for more hostile tribological applications.

Typical Microstructures

For all critical wear applications, it is recommended that both microstructure and hardness values at and close to the surface (where wear is likely to occur) be checked and actually become part of the manufacturing specification. The reasons behind this recommendation are the importance of microstructure in terms of control of engineering properties, as well as the variability that is due to section sensitivity.

Described below are the different microstructures that can be produced in engineering castings. The matrix structure of cast irons usually contains one or more of these constituents: ferrite, pearlite, cementite, phosphide eutectic, martensite, bainitic transformation structures, austenite, and graphite, the latter of which was already been discussed.

Ferrite in cast iron is essentially a single-phase solid solution of silicon in amounts that vary with graphite structure, cooling rate, and silicon content. The amount tends to increase as the cooling rate decreases and the silicon content increases, and the graphite approaches the undercooled form. Fully ferritic structures are normally only obtained by annealing. Ferrite should be avoided in all wear parts.

Pearlite consists of alternate lamellae of ferrite and cementite. The degree of fineness depends on cooling rate. The finer the pearlite, the stronger (harder) is the iron. At low magnification, pearlite appears as a half-tone color. This structure is formed by the transformation of austenite during normal cooling in the mold or in the air through the critical temperature range from 720 to 900 °C (1330 to 1650 °F). This is a highly desirable matrix for sliding wear resistance.

Cementite (Fe_3C) in the massive eutectic form is a hard white constituent formed during solidification, as in mottled or white irons, and in the lamellar form in pearlite, where it is formed by the transformation of austenite through the critical temperature. Eutectic carbides can increase hardness for wear resistance. Harder carbides can be obtained by alloying with Cr and Mo.

Phosphide eutectic (melting point of about 930 °C, or 1706 °F) occurs in two distinct forms in cast irons with more than 0.06% phosphorus. The normal form is pseudobinary, which consists of ferrite and iron phosphide (Fe_3P). The true eutectic forms from the liquidus as austenite plus iron phosphide. Upon cooling, the austenite transforms to ferrite and pearlite, and, with iron phosphide, gives a bulk hardness between 420 and 600 HV.

In mottled irons and in apparently gray irons, this structure can occur in ternary form, where it solidifies as austenite, iron carbide, and iron phosphide. At room temperature after transformation of the austenite, it appears as a eutectic of ferrite, free carbide, and iron phosphide. This form is likely to occur if chromium or vanadium are present in amounts over 0.1% and has a bulk hardness of about 780 to 800 HV, because of the high hardness of the free carbide (800 to 1015 HV). Although it is desirable to improve the sliding wear resistance of gray irons, it markedly increases tool wear in machining operations.

Martensite is a fine acicular, slow-etching structure, normally produced either by very rapid cooling (quenching) of austenite through the critical temperature range or by alloying. The presence of martensite achieves maximum hardness in parts that require abrasion resistance.

Acicular or bainitic transformation structures are produced by either isothermal quenching or alloying. They are often referred to as acicular ferrite and are softer and tougher than martensite, but harder and stronger than pearlite. These acicular structures range from the upper bainites, or acicular ferrites, to martensite, depending on the transformation time, composition, and other factors. Austempered parts provide resistance to the more hostile stress conditions encountered in gear trains, for example.

Austenite can be made stable at room temperature by the addition of alloys, such as nickel and manganese, which depress the critical temperature at which the γ to α change occurs. Although the transformation may have been suppressed at room temperature, it can still occur at a lower temperature, depending on the amount of alloy present. These irons are not normally used for wear resistance.

Applications

Cast irons have been widely used by engineers in applications that require low cost, excellent castability, good damping capacity, ease of machining, and wear resistance. These general terms do not describe precisely the requirements for specific applications, and it is not possible to obtain all of those desirable characteristics in any single alloy.

Cast irons are traditional materials that have an extensive use history and a considerable property database. For tribological purposes, they are the most widely used of all materials. This family of materials sets the standard by which other materials, including composites and ceramics, are compared.

Although there have been many attempts to supersede the use of cast irons in extremely hostile environments, such as in the internal combustion engine, successes are still few in number. The reasons are quite simple and are due to the versatile properties of excellent sliding and abrasive wear resistance, coupled with low cost, casting versatility, dimensional stability over a broad temperature range, and, for certain alloys, ease of machining. Typical wear applications for a variety of cast iron grades are identified in Table 4.

Table 4 Guidelines for industrial applications of selected cast irons for wear parts

Cast iron type	Applications[(a)]
Gray (flake) iron	Piston rings and cylinder bores for compressors and internal combustion engines; machine tool beds and slideways; brake blocks; low-stress gears
Nodular iron	Crankshafts; gears; highly-stressed piston rings
Chilled iron	Cams and tappets where the Hertzian stress is high and lubrication is borderline
White iron (cast and heat treated)	Grinding balls; chute and mill plates and tiles; cone, feeder, deflector, and breaker liner plates; grizzlies
White iron (by hardfacing)	Dragline buckets, shrouds, pick holders, and side plate protection; hardfacing of steel digger teeth for shovels

(a) In all of these applications, there is a wide range of compositions, cooling rates, and heat-treatment variations that control the microstructure and thus the level of wear resistance which can be obtained. A balance also has to be obtained between wear resistance, corrosion resistance, and impact resistance (particularly for mineral handling applications). Great care is therefore required when choosing the optimum material for a specific wear application.

Suppliers and users of cast irons continually update and extend applications into more hostile conditions by improving surface engineering in terms of treatment coatings and machining. A sound engineering understanding of the relationship between composition, manufacture, microstructure, and engineering properties will ensure that these materials continue to be used extensively in tribological applications decades from now.

Friction and Wear of Carbon and Alloy Steels

W.A. Glaeser, Battelle Memorial Institute

Introduction

CARBON STEELS are those steels whose major hardening agent is carbon. Alloying elements such as chromium, vanadium, and molybdenum for improving hardenability are not found in these steels in any significant amount. Carbon steels containing <0.4% C are not hardenable by heat treatment and are not used very much for sliding or rolling contact applications.

Carbon steels have wide and diversified use in tribological applications. Application, rather than specific composition, has determined the need for this class of materials. A few steels have been modified in composition and heat treatment methods for use in bearings. AISI 52100 steel (high-carbon steel), used for ball and roller bearings, is subjected to special vacuum melting practices for inclusion control and is given a special heat treatment to minimize residual austenite content and to ensure dimensional stability. This grade of steel is discussed in the article "Friction and Wear of Bearing Steels" in this Volume.

Austenitic manganese steel (Hadfield manganese steel) is used in the mining, earth moving, and railroad industries where high toughness, impact resistance, and abrasive wear resistance are needed. This alloy contains 1% C and 11 to 14% Mn. It is an austenitic steel that transforms to martensite when abraded heavily. This results in a tough core with a hard skin that continually renews itself as it is worn off. Therefore, this alloy is capable of extensive work hardening under abrasive impact conditions.

Steel is used in machinery for load bearing components such as shafting, gearing, housings, cable, thrust surfaces, and so on. For efficiency, steel surfaces are being coated with wear-resistant materials. In this way, the structural strength of the steel can be used for large components, and a small amount of expensive wear-resistant material is applied to specific areas where it is needed.

Steel Metallurgy

Two abundant elements, iron and carbon, are the basis for the steel industry. Carbon is soluble in iron in small amounts, and steels cover a range of carbon-iron alloys from 0.005 to 2.00 wt% C. The iron-carbon equilibrium diagram, well established from years of research, provides a map of possible phases based on temperature and carbon content. An iron-carbon equilibrium diagram for $\leq$6.67 wt% C is shown in Fig. 1. The portion of the diagram from 0 to 2.0% C covers steels. Higher carbon contents are typical for cast irons.

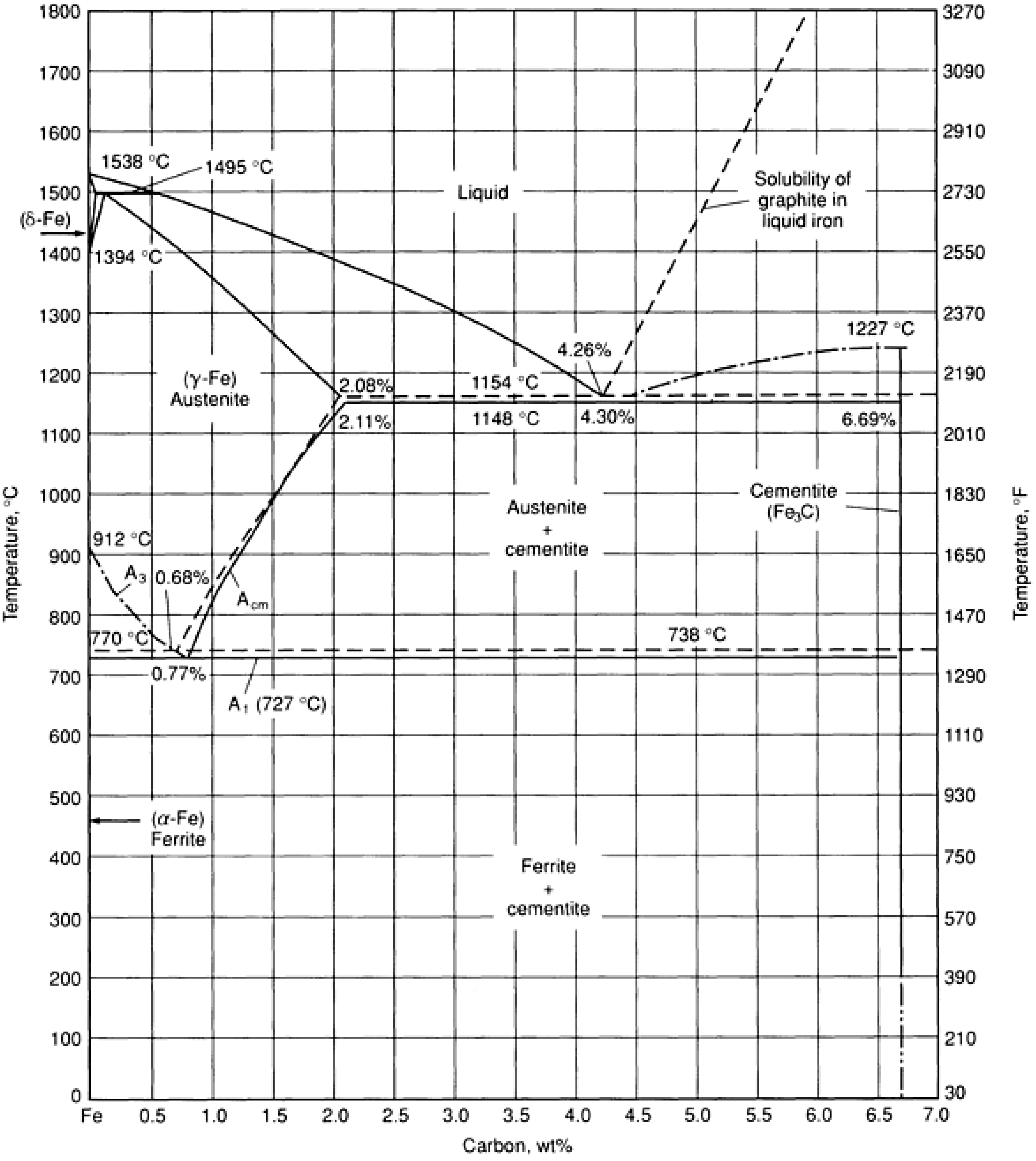


Fig. 1 Iron-carbon equilibrium diagram up to 6.67 wt% C. Solid lines indicate Fe-Fe_3C diagram; dashed lines indicate iron-graphite diagram. Source: Ref 1

Several phases of iron-carbon are shown in Fig. 1: austenite, ferrite, γ-iron, and cementite. These are all equilibrium phases. Austenite exists at elevated temperatures (>727 °C, or 1341 °F). Carbon is an austenite stabilizer and is more soluble in austenite (2.11 wt%) than in ferrite (0.22 wt%). As temperature is reduced from 1148 °C (2098 °F), the solubility of carbon in austenite decreases and carbon is rejected as iron carbide (Fe_3C) or cementite. Two phases (cementite and austenite) exist together as shown in the phase diagram. At 727 °C (1341 °F), austenite converts to ferrite and considerable carbon is rejected from solution as carbide and the two phases (ferrite and cementite) exist together.

Fortunately for engineers, carbon rejection is a very sluggish process and, with rapid enough cooling from the austenite region, a supersaturated solution of carbon in iron results. The lattice strain caused by the excess carbon produces a phase change to body-centered tetragonal (bct), a distorted cubic structure, with very high hardness. This structure, not shown on the equilibrium diagram, is known as martensite. Martensite is desirable as a highly wear-resistant steel phase. Martensite requires rapid cooling such as that achieved by quenching. Slower cooling rates will produce other

microstructures having equally useful engineering properties. These microstructures include bainite (a martensitelike structure) and pearlite (a lamellar structure of ferrite sandwiched between cementite plates). The hardness of these microstructures, especially martensite and bainite, is significantly influenced by carbon content. Figure 2 shows hardness as a function of carbon content for martensite.

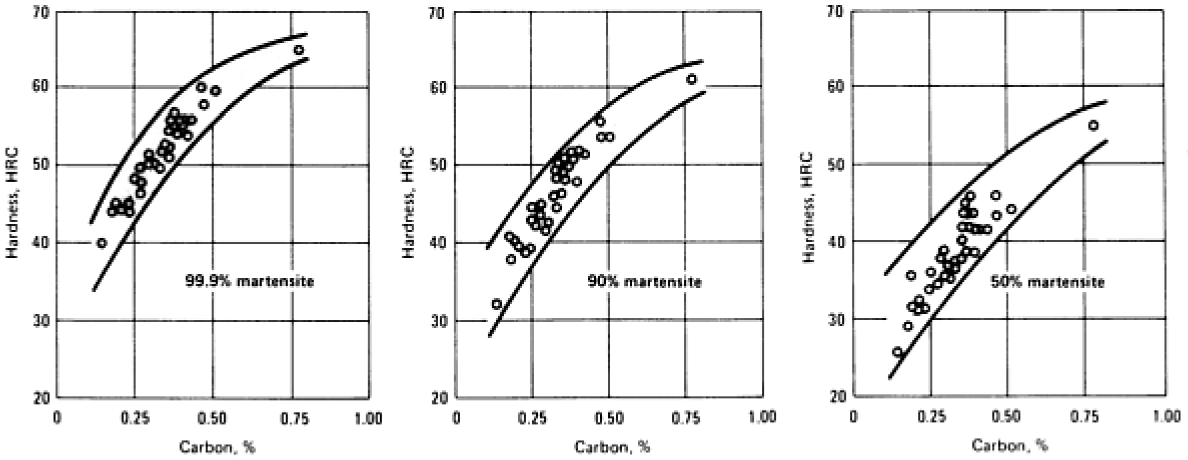


Fig. 2 Plot of hardness versus carbon content for steels containing varying amounts (99.9, 90, and 50%) of untempered martensite in their microstructures

The above metallurgical factors make it possible for a wide variety of steels to be produced by varying the amounts of carbon (Table 1) and by using different heat treatments (Table 2). This makes steel a very versatile engineering material.

Table 1 Nominal compositions and applications of selected abrasion-resistant steel alloys

Material type	Composition, %									Remarks
	C	**Mn**	**Si**	**Cr**	**Ni**	**Mo**	**P**	**S**	**Other**	
Wrought steels										
Mn-Mo-Nb	0.055	1.79	0.15	0.09	0.16	0.31	. . .	. . .	0.20Cu, 0.01N, 0.10Nb, 0.024Ce, 0.01 Al	. . .
Mn-Mo-Nb	0.078	1.33	0.03	0.04	0.09	0.25	. . .	. . .	0.18Cu, 0.033Nb, 0.0054V, 0.003Ti	. . .
Mn-Cr-Ni-Mo-Cu-V	0.11	1.32	0.33	1.44	1.25	0.31	. . .	. . .	0.75Cu, 0.086V	. . .
Mn-V	0.15	1.17	0.56	. . .	. . .	. . .	. . .	. . .	0.12V, 0.10Al	. . .
Mn-Cr-Ni-Mo-Cu-V	0.16	1.44	0.19	1.41	1.21	0.39	0.013	0.007	0.69Cu, 0.06V	. . .
T-1 type A	0.19	0.84	0.28	0.56	. . .	0.18	0.011	0.021	0.02Ti, 0.05V, 0.0032B	. . .
Mn-Cr-Ni-Mo-Cu-V	0.20	1.35	0.32	1.47	1.51	0.40	. . .	. . .	0.76Cu, 0.12V	. . .
Mn-Ni-Mo	0.21	1.71	0.33	. . .	0.73	0.20	. . .	. . .	0.19Cu, 0.01Al	. . .
AISI 1020	0.22	0.45	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
Mn-Ni-Mo	0.24	1.11	0.19	0.04	3.24	0.74	0.021	0.018	. . .	. . .
Mn-Cr-Ni-Mo	0.24	1.18	0.18	0.75	3.14	0.51	0.022	0.019	. . .	. . .

Mn-Cr-Ni-Mo	0.24	1.22	0.20	1.43	3.15	0.25	0.022	0.019	. . .	. . .
Mn-Cr-Ni-Mo	0.25	1.17	0.30	1.39	0.87	0.76	0.022	0.019	. . .	. . .
Mn-Cr-Ni-Mo	0.25	1.22	0.28	1.40	1.58	0.47	0.022	0.019	. . .	. . .
Mn-Ni-Mo-Cu-B	0.27	1.90	0.23	. . .	0.74	0.17	. . .	. . .	0.29Cu, 0.014Al, 0.0016B, 0.04Ti	. . .
Mn-Ni-Mo-B	0.27	2.10	0.30	. . .	0.82	0.17	. . .	. . .	0.16Cu, 0.06Al, 0.0017B, 0.04Ti	. . .
Mn-Mo-Cu-B	0.28	1.60	0.25	0.05	0.06	0.16	0.011	0.025	0.28Cu, 0.023Ti, 0.049Al, 0.002B, 0.004N	. . .
Mn-Ni-Mo-B	0.29	2.36	0.24	. . .	0.86	0.15	. . .	. . .	0.08Cu, 0.015Al, 0.0017B, 0.04Ti	. . .
Cr-Mo-B	0.30	0.82	0.22	0.35	0.04	0.27	0.010	0.017	0.03Cu, 0.026Ti, 0.039Al, 0.001B, 0.006N	. . .
Cast lean manganese-austenitic steels										
6Mn-5Cr-1Mo	0.7	6.02	0.53	5.07	. . .	1.01	0.019	0.015	. . .	. . .
6Mn-5Cr-1Mo	1.05	6.12	0.54	5.04	. . .	1.00	0.021	0.019	. . .	. . .
6Mn-5Cr-1Mo	1.22	6.02	0.55	4.96	. . .	0.95	0.020	0.019	. . .	. . .
9Mn-1Mo-Ti	1.10	9.52	0.59	. . .	. . .	1.07	. . .	. . .	0.36Ti	. . .
9Mn-1Mo-Ti	1.21	9.13	0.82	. . .	. . .	0.99	. . .	. . .	0.134Ti	. . .
9Mn-1Mo-Ti	1.27	9.49	0.56	. . .	. . .	1.08	. . .	. . .	0.32Ti	. . .
9Mn-1Mo-Ti	1.30	8.61	0.50	. . .	. . .	0.93	. . .	. . .	0.27Ti	. . .
9Mn-1Mo-Ti	1.36	8.47	0.49	. . .	. . .	1.00	. . .	. . .	0.064Ti	. . .
Cast austenitic-manganese steels										
12Mn-1Mo	0.65	12.74	0.51	. . .	. . .	0.96	. . .	. . .	. . .	. . .
12Mn	0.93	12.97	0.50	. . .	. . .	. . .	. . .	. . .	. . .	. . .
12Mn-1Mo	0.93	12.0	0.5	. . .	. . .	1.0	. . .	. . .	. . .	. . .
12Mn-1Mo	0.97	12.5	0.5	. . .	. . .	0.94	. . .	. . .	. . .	. . .
12Mn	1.1	12.5	0.5	. . .	. . .	. . .	. . .	. . .	. . .	. . .
12Mn	1.24	12.5	0.5	. . .	. . .	0.05	. . .	. . .	. . .	. . .
12Mn-1Mo-Ti	1.26	12.5	0.5	. . .	. . .	0.96	. . .	. . .	0.25Ti	. . .
12Mn-1Mo-Ti	1.29	12.5	0.5	. . .	. . .	0.94	. . .	. . .	0.018Ti	. . .
12Mn-1Mo-Ti	1.29	12.5	0.5	. . .	. . .	1.02	. . .	. . .	0.13Ti	. . .
12Mn-1Mo-Ti	1.31	12.5	0.5	. . .	. . .	0.92	. . .	. . .	. . .	. . .
Cr-Ni-Mo	0.27	. . .	. . .	1.0	3.0	2.0	. . .	. . .	. . .	. . .
AISI 4340	0.4	0.7	0.3	0.8	1.8	0.25	. . .	. . .	. . .	. . .
Mn-Cr-	0.65	1.75	0.13	0.75	. . .	0.13	. . .	. . .	. . .	. . .

Mo										
C-Mn	0.80	1.00	0.18	. . .	. . .	. . .	. . .	. . .	. . .	. . .
Cast steels										
Si-Cr-Mo	0.34	0.84	1.92	1.92	0.22	0.57	. . .	. . .	. . .	. . .
Mn-Si-Cr-Mo	0.43	1.39	1.46	0.83	. . .	0.49	. . .	. . .	. . .	. . .
Mn-Si-Cr-Ni-Mo	0.55	1.44	1.32	0.68	0.96	0.63	. . .	. . .	0.08Cu	. . .
Mn-Si-Cr-Mo	0.63	1.44	1.48	0.83	. . .	0.49	. . .	. . .	. . .	. . .
Cr-Mo	0.63	0.71	0.58	2.30	. . .	0.34	0.028	. . .	. . .	. . .
Cr-Mo	0.88	0.95	0.72	2.44	. . .	0.35	0.027	. . .	. . .	. . .
Carburizing steels										
1015	0.13-0.18	0.30-0.60	. . .	. . .	. . .	. . .	. . .	. . .	. . .	Little used; mainly for small thin parts such as needle rollers
1019	0.15-0.20	0.70-1.00	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
1020	0.18-0.23	0.30-0.60	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
1118	0.14-0.20	1.30-1.60	. . .	. . .	. . .	. . .	. . .	. . .	. . .	Free-machining carburizing grade resulfurized
4023	0.20-0.25	0.70-0.90	0.15-0.30	. . .	. . .	0.20-0.30	. . .	. . .	. . .	. . .
4027	0.25-0.30	0.70-0.90	0.20-0.35	. . .	. . .	0.20-0.30	. . .	. . .	. . .	Moderate strength and toughness
4022	0.20-0.25	0.70-0.90	0.20-0.35	. . .	. . .	0.35-0.45	. . .	. . .	. . .	. . .
5120	0.17-0.22	0.70-0.90	0.15-0.30	0.70-0.90	. . .	. . .	. . .	. . .	. . .	. . .
4118	0.18-0.23	0.70-0.90	0.15-0.30	0.40-0.60	. . .	0.08-0.15	. . .	. . .	. . .	. . .
4720	0.17-0.22	0.50-0.70	0.35-0.55	0.35-0.55	0.90-1.20	0.15-0.25	. . .	. . .	. . .	Cr and Ni produce increased hardenability for heavy sections (large roller bearings)
4820	0.18-0.23	0.50-0.70	0.15-0.30	. . .	3.25-3.75	0.20-0.30	. . .	. . .	. . .	. . .
4320	0.17-0.22	0.45-0.65	0.15-0.30	0.40-0.60	1.65-2.00	0.20-0.30	. . .	. . .	. . .	. . .
8620	0.18-0.23	0.70-0.90	0.15-0.30	0.40-0.60	0.40-0.70	0.15-0.25	. . .	. . .	. . .	. . .
9310	0.08-0.13	0.45-0.65	0.15-0.30	1.00-1.40	3.00-3.50	0.08-0.15	. . .	. . .	. . .	For very high shock resistance; high fatigue resistance; surface carbon content must not exceed 0.9% to avoid retained austenite
3310	0.08-0.13	0.45-0.60	. . .	1.40-1.75	3.25-3.75	. . .	. . .	. . .	. . .	. . .

Table 2 Abrasive wear data for selected steels

Type	Heat treatment	Hardness, HB	Gouging wear ratio	Rubber wheel weight loss, g
AISI 4140 (0.40 C)	845 °C (1550 °F), oil quenched	582	. . .	0.219
AISI 4140 (0.40 C)	845 °C (1550 °F), oil quenched, 540 °C (1000 °F)	550	. . .	0.235
AISI 4140 (0.40 C)	845 °C (1550 °F), oil quenched, 540 °C (1000 °F)	499	. . .	0.410
AISI 4140 (0.40 C)	845 °C (1550 °F), oil quenched, 540 °C (1000 °F)	363	. . .	0.531
AISI 4340 (0.40 C)	Normalized	320	0.674	. . .
AISI 4340 (0.40 C)	Quenched 650 °C (1200 °F)	340	0.716	. . .
AISI 4340 (0.40 C)	Quenched 205 °C (400 °F)	520	0.232	. . .
AISI 1085 (0.82 C)	Hardened and tempered	456	. . .	0.281
AISI 1090 (0.95 C)	Hardened and tempered	450	. . .	0.278
AISI 1090 (0.98 C)	Hardened and tempered	455	. . .	0.220
6Cr-1Mo (0.88 C)	1065 °C (1950 °F), air cooled, 230 °C (450 °F)	601	0.112	. . .
6Cr-1Mo (0.88 C)	1065 °C (1950 °F), air cooled, 540 °C (1000 °F)	601	0.148	. . .
5Mn-1Mo-2Cr (1.0 C)	1040 °C (1900 °F), furnace cooled	288	0.245	. . .

6Mn-1Mo (1.27 C)	1040 °C (1900 °F), water quenched, 60 °C (140 °F)	200	0.192	. . .
6Mn-1Mo, Ti (1.23 C)	1040 °C (1900 °F), water quenched, 60 °C (140 °F)	200	0.170	. . .
6.5Mn-1Mo (1.01 C)	1040 °C (1900 °F), furnace cooled	240	0.292	. . .
6.5Mn-1Mo-2Cr (0.99 C)	1040 °C (1900 °F), furnace cooled	241	0.316	. . .
6.5Mn-1Mo-5Cr (1.0 C)	1040 °C (1900 °F), furnace cooled	246	0.324	. . .
6.5Mn-3Mo-2Cr (1.02 C)	1040 °C (1900 °F), furnace cooled	241	0.294	. . .
6.5Mn-2Cr (1.00 C)	1040 °C (1900 °F), furnace cooled	474	0.329	. . .
8Mn-1Mo-1Cr (1.00 C)	1040 °C (1900 °F), furnace cooled	229	0.337	. . .
9Mn-1Mo (1.27 C)	1065 °C (1950 °F), water quenched, 60 °C (140 °F)	206	0.219	. . .
9Mn-1Mo, Ti (1.24 C)	1065 °C (1950 °F), water quenched, 60 °C (140 °F)	199	0.213	. . .
12Mn (0.93 C)	1040 °C (1900 °F), water quenched	185	0.328	. . .
12Mn (1.00 C)	1040 °C (1900 °F), water quenched	199	0.279	. . .
12Mn (1.24 C)	1040 °C (1900 °F), water quenched	198	0.212	. . .
12Mn (1.27 C)	1065 °C (1950 °F), water quenched, 60 °C (140 °F)	211	0.207	. . .

Not only is there a large variety of steels from which to choose, but within those grades of steels, various heat treatments are possible, resulting in a greater variety of microstructures. The microstructure of steel can have a significant effect on wear resistance (Table 3). Microstructure can also influence corrosion resistance, toughness, dimensional stability, residual stresses, and fatigue resistance.

Table 3 Hardness, toughness, and abrasive wear data for abrasion-resistant steel alloys

Material	Hardness, HB	Charpy V-notch impact energy		Gouging wear ratio[a]	Weight loss, g		Microstructure[b]
		J	**ft · lbf**		**Pin**	**Rubber wheel**	
Wrought steels							
Mn-Mo-Nb (0.06C)	212	146	108	1.26	0.1429	. . .	F
Mn-Mo-Nb (0.06C)	228	146	108	1.12	0.1397	. . .	F
Mn-Mo-Nb (0.08C)	187	106	78	1.51	0.1560	. . .	F
Mn-Cr-Ni-Mo-Cu-V (0.11C)	290	23	17	0.88	0.1240	. . .	B,M
Mn-Cr-Ni-Mo-Cu-V (0.11C)	290	. . .	. . .	0.80	0.1248	. . .	B,M
Mn-V (0.15C)	225	85	63[c]	1.20	. . .	. . .	F,p
Mn-Cr-Ni-Mo-Cu-V (0.16C)	333	. . .	. . .	0.75	0.1179	. . .	M,B
T-1 type A (0.19C)	260	57	42	1.08	0.1353	0.845	M,B
T-1 type A (0.19C)	376	. . .	. . .	0.64	0.1240	. . .	F,c
Mn-Cr-Ni-Mo-Cu-V (0.20C)	380	69	51	0.48	0.1123	. . .	B,M
Mn-Cr-Ni-Mo-Cu-V (0.20C)	380	. . .	. . .	0.50	0.1123	. . .	B,M
Mn-Ni-Mo (0.21C)	444	. . .	. . .	0.39	0.1063	0.496	M,B
AISI 1020 (0.22C)	106	163	120	1.87	0.156	. . .	F,p
Mn-Ni-Mo (0.24C)	251	60	44	0.84	0.1290	. . .	F,c
Mn-Cr-Ni-Mo (0.24C)	263	47	35	0.79	0.1265	. . .	F,c
Mn-Cr-Ni-Mo (0.25C)	273	46	34	0.77	0.1257	. . .	F,c
Mn-Cr-Ni-Mo (0.25C)	294	41	30	0.66	0.1199	. . .	F,c
Mn-Cr-Ni-Mo (0.25C)	286	47	35	0.71	0.1226	. . .	F,c
Mn-Ni-Mo-Cu-B (0.27C)	333	. . .	. . .	0.69	0.1215	0.713	M,B
Mn-Ni-Mo-B (0.27C)	394	. . .	. . .	0.67	0.1211	0.662	M,B
Mn-Mo-Cu-B (0.28C)	379	. . .	. . .	0.72	0.1216	0.630	M,B
Mn-Mo-Cu-B (0.28C)	434	. . .	. . .	0.52	0.1162	0.550	M,B
Mn-Ni-Mo-B (0.29C)	462	. . .	. . .	0.47	0.1088	0.512	M,B
Cr-Mo-B (0.30C)	395	. . .	. . .	0.59	0.1156	0.520	M,B
Cr-Mo-B (0.30C)	456	. . .	. . .	0.44	0.1108	0.440	M,B
Cr-Ni-Mo (0.27C)	510	. . .	. . .	0.36	0.0954	0.756	M
Cr-Ni-Mo (0.27C)	520	. . .	. . .	0.46	0.0964	0.738	M
AISI 4340 (0.4C)	320	8.1	6.0	0.67	. . .	. . .	F,P
AISI 4340 (0.4C)	340	29.8	22.0	0.72	. . .	. . .	F,c
AISI 4340 (0.4C)	520	10.6	7.8	0.23	. . .	. . .	M
Mn-Cr-Mo (0.65C)	600	. . .	. . .	. . .	. . .	0.0760	M
C-Mn (0.80C)	292	. . .	. . .	. . .	. . .	. . .	P
Cast lean manganese-austenitic steels							
6Mn-5Cr-1Mo (0.7C)	339	6.8	5.0	0.30	. . .	. . .	A,c
6Mn-5Cr-1Mo (0.7C)	243	25	18.3	0.34	. . .	. . .	A,c

6Mn-5Cr-1Mo (1.0C)	352	5.4	4.0	0.27	. . .	. . .	A,c
6Mn-5Cr-1Mo (1.0C)	285	12	9.0	0.29	. . .	. . .	A,c
6Mn-5Cr-1Mo (1.2C)	385	4.7	3.5	0.27	. . .	. . .	A,c
6Mn-5Cr-1Mo (1.2C)	275	7.9	5.8	0.28	. . .	. . .	A,c
9Mn-1Mo-Ti (1.1C)	207	41	30	0.34	. . .	. . .	A
9Mn-1Mo-Ti (1.2C)	224	20	14.7	0.21	. . .	. . .	A
9Mn-1Mo-Ti (1.25C)	207	65	48	0.27	. . .	. . .	A
9Mn-1Mo-Ti (1.3C)	. . .	24	18	0.25	. . .	. . .	A
9Mn-1Mo-Ti (1.35C)	208	22	16.3	0.18	. . .	. . .	A
Cast austenitic-manganese steels							
12Mn-1Mo (0.65C)	191	119	88	0.42	. . .	. . .	A
12Mn (0.95C)	185	138	102	0.33	. . .	. . .	A
12Mn-1Mo (0.95C)	188	72	53	0.32	0.0871	. . .	A,p
12Mn-1Mo (1.09C)	192	145	107	0.29	. . .	. . .	A
12Mn (1.1C)	199	. . .	. . .	0.28	0.0821	. . .	A
12Mn (1.25C)	198	80	59	0.21	. . .	. . .	A
12Mn-1Mo-Ti (1.25C)	201	72	53	0.21	. . .	. . .	A
12Mn-1Mo-Ti (1.3C)	204	87	64	0.22	. . .	. . .	A
12Mn-1Mo-Ti (1.3C)	201	77	57	0.22	. . .	. . .	A
12Mn-1Mo-Ti (1.3C)	199	34	25	0.21	. . .	. . .	A

(a) Weight loss of test specimen divided by the weight loss of reference specimen (T-1 type A steel, 260 HB).

(b) A, austenite; M, martensite; F, ferrite; B, bainite; C and c, carbide; P and p, pearlite; lower case letters indicate small amounts.

(c) $\frac{2}{3}$ size impact specimen

Steel Transformation Diagram

When carbon steel is heat treated by quenching from a high temperature, the resulting room-temperature microstructures are not those shown in the equilibrium diagram. Instead, a temperature-time-transformation diagram (TTT) diagram is used to describe these changes. A typical diagram for AISI 1095 steel is shown in Fig. 3. The left side of the diagram shows the temperature in both centigrade and fahrenheit. The right side of the diagram shows the hardness of the steel at room temperature. Note that the maximum hardness attainable for this steel is 66 HRC. The first heavy curved line in the chart represents the time at which transformation begins. In hypoeutectoid steels, this initial transformation consists of the separation of proeutectoid ferrite. This is followed by the separation of ferrite and carbide in the form of pearlite. The beginning of the pearlite formation is represented by the second heavy curved line. Below the knee of the curve ($\leq$540 °C, or 1000 °F), the transformation product is bainite, a phase similar in morphology to martensite but not quite as hard. If the solid curved line is not crossed before the martensite start temperature (M_s) horizontal dashed line, martensite is the product of austenite decomposition. This diagram shows that a very rapid quench is required to produce martensite (that is, one must go from 885 °C, or 1625 °F, past the knee in <1 s--a quench rate of approximately 360 °C/s, or 650 °F/s). Bainite can be obtained by rapid quenching to a temperature just below the knee and holding at that temperature until the transformation is complete, followed by a quench to room temperature.

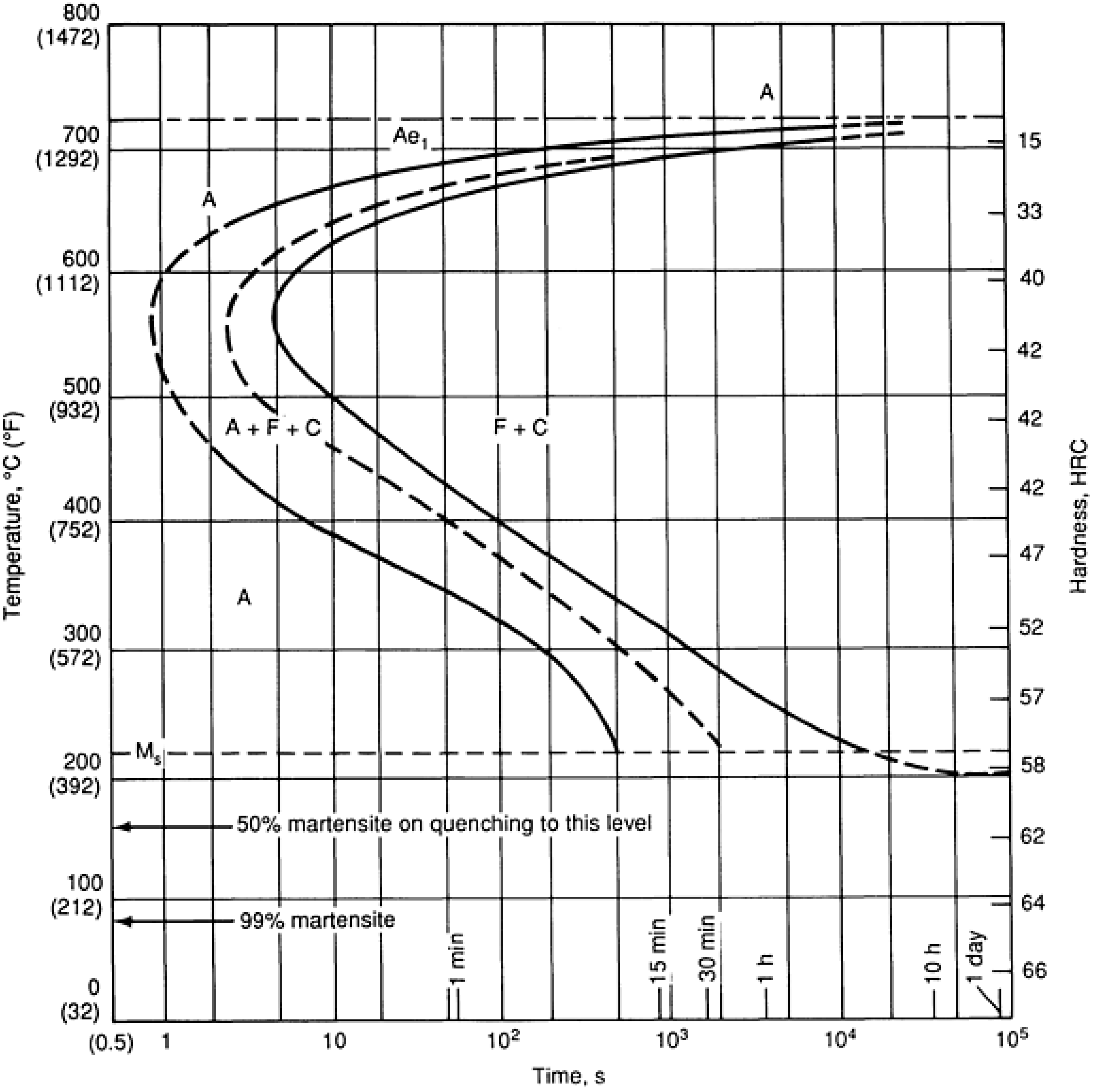


Fig. 3 Time-temperature-transformation diagram for SAE 1095 eutectoid carbon steel (0.9C-0.3 Mn). Austenitized at 885 °C (1625 °F); grain size, 4 to 5. Ae_1, equilibrium transformation temperature

A variety of microstructures can be obtained by controlling the quench processes. Martensite is the hardest phase and provides the highest wear resistance in the absence of any fracture-initiating conditions. If a tougher material is desired, bainite is the best wear-resistant phase. Of course, tempering (that is, heating the martensite to an elevated temperature below the austenitizing temperature and holding to reduce the lattice strain) is a process that can be used to increase the ductility and toughness of martensite. The maximum martensitic hardness attainable in a steel is a function of the carbon content (Fig. 2).

Wear Properties of Carbon Steel

Hardness as a Function of Carbon Content. Under the abrasive conditions found in mining and construction operations, wear rates of steel can be related to hardness and carbon content. For example, wear tests were conducted on steel rods used on a vibrating screen at an ore crushing plant. These screens were equipped with rods 6.5 mm ($\frac{1}{4}$ in.) or 4.8 mm ($\frac{3}{16}$ in.) in diameter, 585 mm (23 in.) long, and equally spaced in several rows across the rod deck of vibrating screens used to size -25 mm (-1 in.) siliceous ore into ±9.5 mm (±$\frac{3}{8}$ in.) sizes. The wear rate of all test rods was compared with that of 1070 high-carbon steel that was oil quenched and tempered to 44 HRC. To ensure that each type of

steel would be exposed to the same abrasion, rods were alternated in groups of five across the screen. Wear rates based on the loss of weight for the different steels and hardnesses are shown in Fig. 4. Although the 0.30C-13Cr stainless steel showed the best resistance to wear, 1080 steel austempered to 57 HRC was found to be more cost effective.

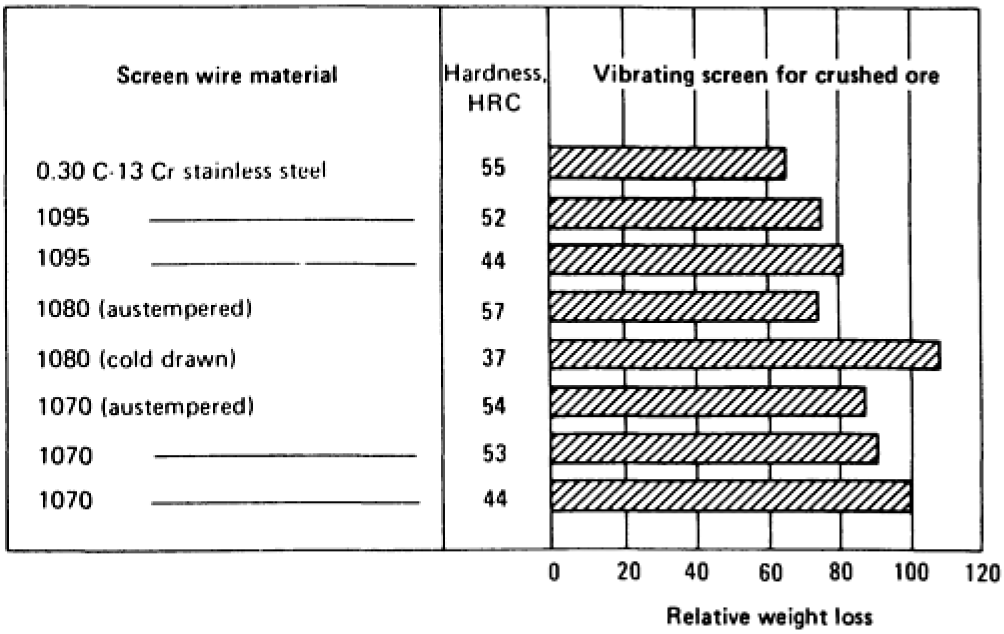


Fig. 4 Effect of hardness and carbon content on the abrasive wear of selected steel rods used in an ore crusher

Improving Wear Properties of Mild Steels. Mild steel demonstrates poor wear resistance and resistance to surface damage during dry sliding. The use of mild steel in sliding surface contact requires surface treatment, such as hardening or coating, and the selection of a "compatible" mating material such as bronze or babbitt. Where hard minerals come in contact with steel, wear is very rapid unless the steel surface is hardened or coated with a very hard material.

Corrosion Resistance Improved by Altering Microstructure. Steel is subject to accelerated wear in a corrosive environment. Unprotected steel is also susceptible to fretting damage or the formation of oxidized wear debris between two contacting surfaces in low-amplitude oscillating motion. A wide variety of microstructures is possible in the heat treatment of steel or cast iron. Wear properties can be related to specific microstructures.

Steel Selection Based on Relative Costs

When selecting a steel based on its wear-resistance properties, the total cost of the steel and its heat treatment must be considered. The following steels, which may have suitable wear-resistance properties in specific applications, are listed in order of increasing total costs:

- Low-carbon steels, such as 1020, not heat treated
- Simple high-carbon steels, such as 1095, not heat treated
- Directly hardened carbon or low-alloy steels, either through-hardened or surface-hardened by induction or flame process
- Low-carbon or low-alloy steels that are surface-hardened by carburizing, cyaniding, or carbonitriding
- Medium-carbon chromium or chromium-aluminum steels that are hardened by nitriding
- Directly hardened high-alloy steels, such as D2 high-carbon high-chromium tool steel (1.50C-12Cr), that contain particles of free carbide
- Precipitation-hardening stainless steels (mainly for applications involving elevated temperatures and corrosive environments, as well as excessive wear)
- Specialty steels produced by powder metallurgy (P/M) or mechanical alloying techniques

- Alloy carbides bonded by steel matrices

Other ferrous materials, such as high-manganese austenitic steels and various classes of cast irons, are also widely used for wear-resistance applications.

Depth of Hardened Regions

Skids, grinding rods, chute liners, and similar parts may be considerably reduced in section before replacement is necessary. In such parts, a more expensive deep-hardening steel may be more economical than a shallow-hardening steel. For example, a 64 mm ($2\frac{1}{2}$ in.) diameter bar with a surface hardened of 50 HRC may be made of either a water-quenched 1040 or an oil-quenched 5160 steel. However, by the time the bar has been worn to three-fourths of its original diameter (~48 mm, or $1\frac{7}{8}$ in.), the 1040 steel will have a surface hardness of ~25 HRC and thus would wear at a much faster rate than the 5160 steel, which has a hardness of ~37 HRC at the same location (Ref 2).

Toughness

Wear resistance tends to increase with hardness, but it decreases as toughness increases. This is an important relationship in applications that require both wear resistance and impact resistance.

The correlation between wear resistance and toughness for a variety of ferrous alloys is shown in Fig. 5. The scatter arises, at least in part, from microstructural effects. For example, point 22B refers to AISI 4340 steel, quenched and tempered at a high temperature of ~650 °C (~1200 °F) to produce fine carbides in a ferrite matrix. Point 22A represents the same steel, except normalized to produce fine pearlite; point 22C represents a quenched sample tempered at 205 °C (400 °F), a relatively low tempering temperature. Steels in the lower band of Fig. 5 combine toughness with wear resistance; these are mainly the austenitic manganese steels. The data in Fig. 5 indicate that, for most ferrous alloys, there is a trade-off between wear resistance and toughness. In some alloys, altering the carbon content is a simple method for adjusting these properties.

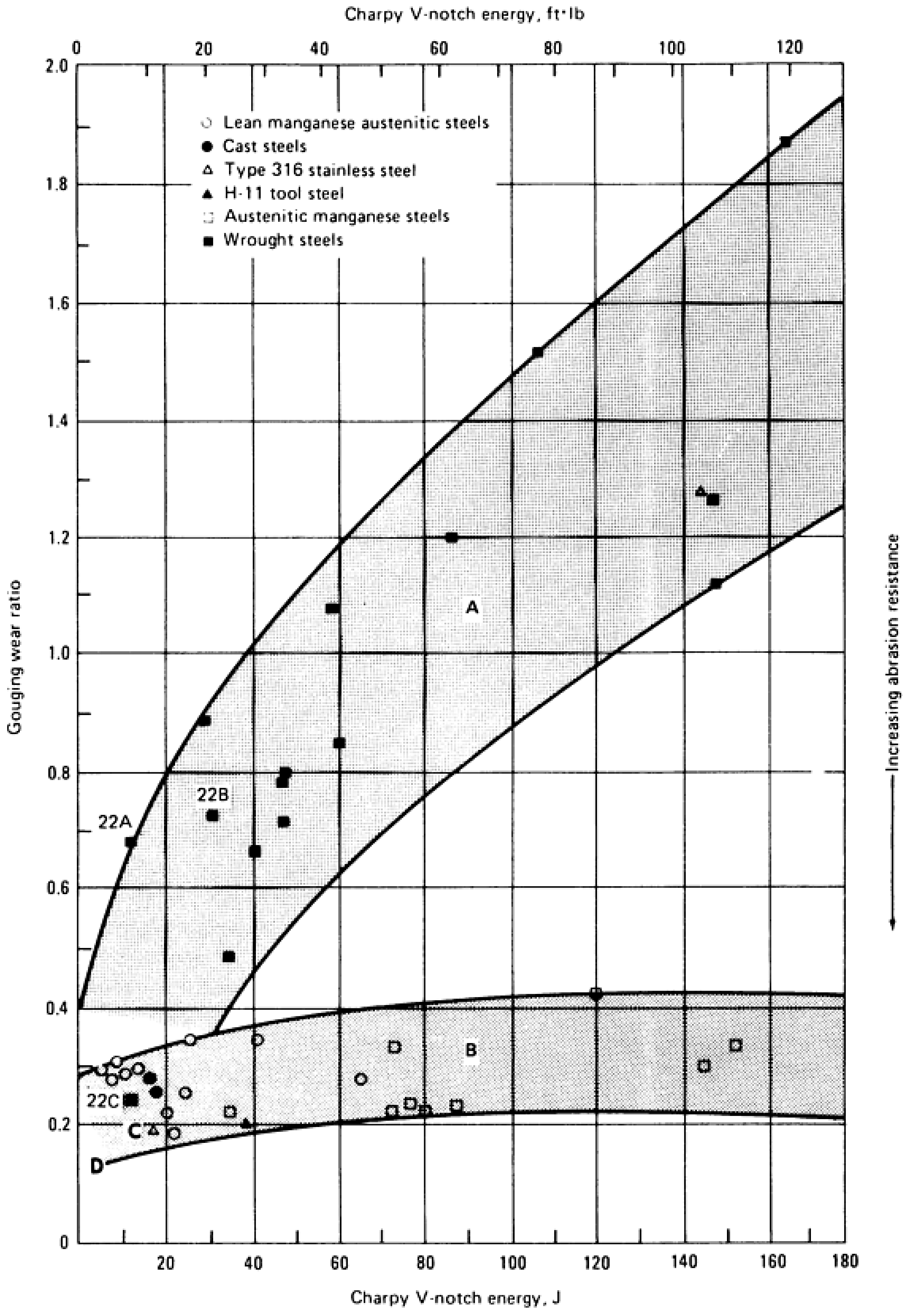


Fig. 5 Relationship between resistance to gouging abrasion and toughness of selected materials. Area A, wrought and cast low-alloy steels; area B, austenitic manganese steels; area C, variety of heat-treated steels; area D, high-chromium white cast irons. Source: Ref 3

Carbon Content

The wear resistance of ferritic steel is improved by hardening, either throughout the section or superficially. The maximum hardness depends on the carbon content of the steel and the amount of martensite (that is, efficiency of quenching) (Fig. 2)

Standard hardness measurements may indicate that a martensitic steel is largely transformed, although it may retain some austenite. Exposure to ultralow temperatures, followed by tempering, can help complete the transformation to martensite and improve wear resistance. Because martensite is a metastable structure, it begins to transform to more stable structures as the temperature is increased. Consequently, martensitic steels are not suitable for wear resistance at elevated temperatures (>200 °C, or 390 °F) or for applications in which the heat of friction can raise the temperature significantly. Special alloy steels, such as tool steels or martensitic stainless steels, are appropriate for service at higher temperatures. The thermal instability of martensite should also be considered during finishing operations (such as grinding) when a heat-affected zone (HAZ) could be produced at the surface. The resultant tempering effects could be localized or general; in either case, wear resistance is likely to be reduced.

Carbon content also affects hardness and wear resistance through the formation of various simple and complex carbides. Wear properties depend on the type, amount, shape, size, and distribution of carbides present, in addition to the properties of the matrix (for example, hardness, toughness, and stability). Despite this complexity, a correlation for relative wear content is possible.

Relation of Hardness to Microstructure

The frequent use of bulk hardness as a guide to abrasive wear resistance is supported by the data shown in Fig. 6 for annealed unalloyed metals. These data were obtained using an abrasive cloth (two-body abrasion) with an abrasive hardness much greater than that of the metal samples. The data points are approximate; the experimental scatter of the measurements is not shown.

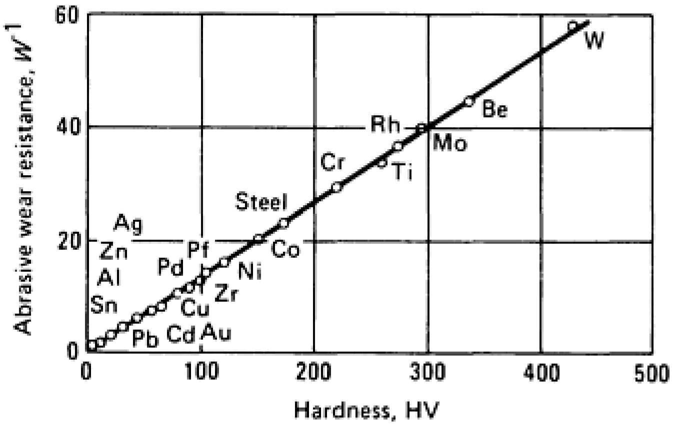


Fig. 6 Abrasive wear resistance versus hardness for annealed unalloyed metals and steel. Source: Ref 4

Corresponding correlations with other properties related to hardness (such as elastic modulus) have also been presented. In all cases, if the metals are unalloyed, a simple correlation is obtained for controlled tests of two-body abrasion. Different crystal structures would be expected to yield different correlations, but the data in Fig. 6 do not show such an effect.

Care must be exercised in extending the simple hardness correlation to metals containing impurities or solutes, or to more complicated alloys. Figure 7 shows how wear resistance relates to hardness for various types of materials. The linear plot shown for pure metals in Fig. 6 is repeated as the steep line in Fig. 7. Another straight line describes brittle ceramics reasonably well. The differences in bonding type may account for these two distinct lines.

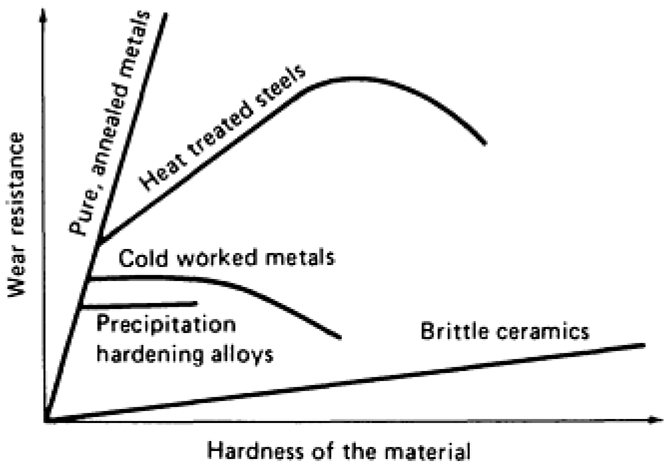


Fig. 7 Plot of wear resistance versus hardness for selected materials (metals, alloys, and ceramics) having various microstructures. Source: Ref 5

Between the curves for pure metals and brittle ceramics are curves for heat-treated steels, cold-worked metals, and precipitation-hardening alloys of varying hardnesses. For a typical precipitation-hardening alloy, large increases in hardness provide very little improvement in wear resistance. The same is true of simple cold working alloys (assuming no complicating phase changes), except that wear resistance actually decreases at high hardness levels. Heat-treated steels show initial improvement in wear properties as hardness increases, but eventually the correlation is reversed. It is possible that the reduced wear resistance at high hardness for cold-worked metals and heat-treated steels is related to the reduction in fracture toughness. Zum Gahr (Ref 6) has evaluated the effect of low fracture toughness on reduced abrasion resistance of metal alloys. He suggests that during abrasion a critical load can be reached by a given abrasive particle, above which fracture toughness becomes a significant factor in the abrasive wear process. Although carbon steels are not likely to fall into the category of materials for which fracture toughness related abrasive wear is a problem, the effect of heat treatment and carbon content on fracture toughness should be kept in mind. As the carbon content of carbon steel increases, the maximum martensitic hardness attainable by quenching increases and fracture toughness decreases. If one tries to use very hard untempered martensite in a part, the chances of the microfracture dominating wear in abrasive wear increases.

Pearlitic steels are inexpensive materials and above a certain hardness level exhibit reasonable wear resistance. Clayton (Ref 7) demonstrated in pin-on-ring experiments that a hardness of >250 HV 30 (Vickers hardness of >250 measured under a load of 30 kgf applied for 10 to 15 s, the normal loading time) wear resistance increased by almost an order of magnitude (Fig. 8). Note that specific wear, which is defined as the volume removed per total sliding distance, has units of mm^3/mm. Wear resistance, which is the reciprocal of specific wear, has units of mm/mm^3 or mm^{-2}.

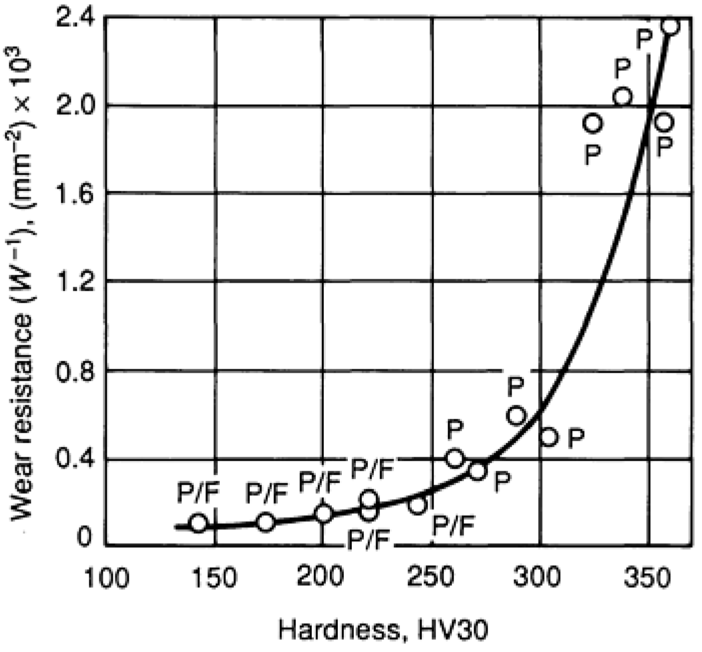


Fig. 8 Dry pin-on-ring test used to evaluate effect of microstructure on the wear resistance and the hardness of pearlitic steel pins. Test parameters: pressure, *P,* 34.71 MPa (5.033 ksi); velocity, *v,* 0.4 m/s (1.3 ft/s); atmosphere, dry air. P, pearlite; P/F, pearlite and ferrite. Source: Ref 6

In lubricated block-on-ring wear experiments with pearlitic carbon steel where deformation wear was the principal wear mode, pearlite was observed to minimize the depth of extreme plastic deformation (Ref 8). During heavy working, the interlath spacing in pearlite will decrease and the ferrite will work harden. Thus, the hardness of pearlite increases with decreased lamellae spacing (Fig. 9) and the structure tends to resist recrystallization (Ref 9). The result is a thinner zone of heavy working and a smaller volume of metal extruding out of the contact zone.

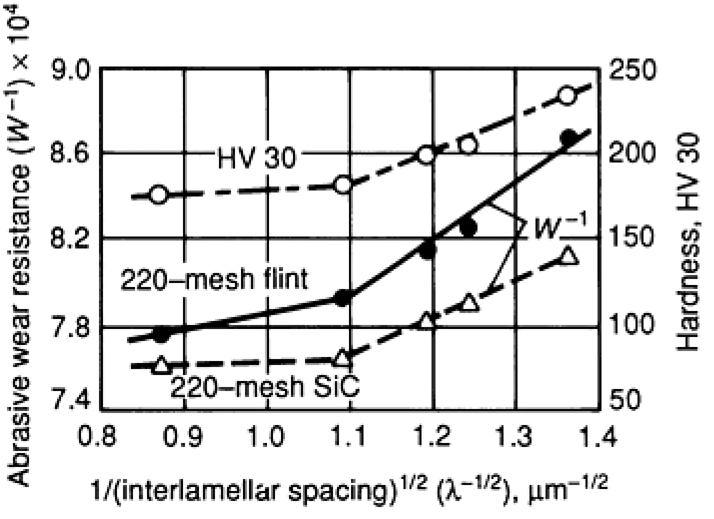


Fig. 9 Abrasive wear resistance and bulk hardness of 0.7% C steel as a function of pearlite interlamellar spacing. Abrasion data obtained using pin abrasion test apparatus with pressure, *P,* of 710 kPa (105 psi); two different 220-mesh abrasives (flint [96 to 99% Si] and silicon carbide) used in separate tests. Source: Ref 6

Under boundary lubrication conditions, where heavier loads are possible than under dry conditions, a pearlitic structure with tight interlamellar spacing in the pearlite phase is appropriate to minimize the depth of shear instability. Shear instability (Ref 10) below the contact surface is the source for deformation wear.

Austenitic steels with varying manganese content show wear resistance to be a function of manganese content. The resistance depends on the transformation of the unstable austenite to martensite (that is, the hard wear-resistant phase). Manganese tends to promote the retention of room-temperature austenite in carbon steel. The retained austenite is unstable and will transform to martensite under heavy deformation. The room-temperature austenite becomes more stable as the manganese content of the alloy is increased. Pin-on-disk experiments performed by Jost and Schmidt (Ref 11) showed that the amount of austenite transformation decreased with increasing manganese content. Figure 10 shows an almost linear increase in wear resistance with decreasing concentration of manganese in the alloy.

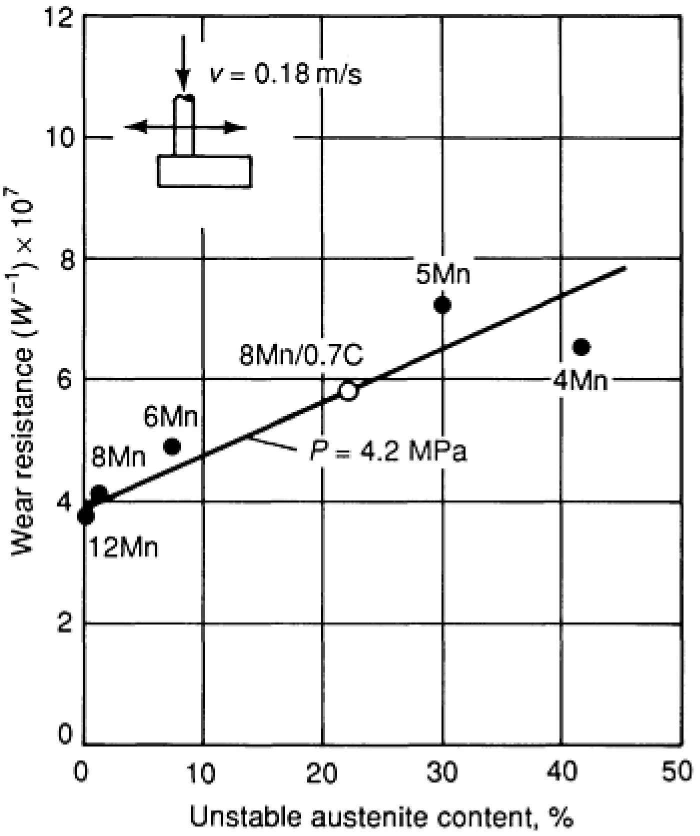


Fig. 10 Plot of wear resistance versus unstable austenite content as a function of manganese content. Pin-on-disk test specimens had 4 to 8% Mn content Fe-Mn-C steel pins rubbed against a steel disk. Test parameters, pressure, 4.2 MPa (610 psi); velocity, 0.18 m/s (0.59 ft/s). Source: Ref 6

References

1. G. Krauss, Physical Metallurgy and Heat Treatment of Steel, *Metals Handbook Desk Edition,* H.E. Boyer and T.L. Gall, Ed., American Society for Metals, 1985, p 28-2 to 28-10
2. *Properties and Selection of Irons and Steels,* Vol 1, 9th ed., *Metals Handbook,* 1978, p 606
3. D.E. Diesburg and F. Borik, Optimizing Abrasion Resistance and Toughness in Steels for the Mining Industry, *Symposium on Material for the Mining Industry,* Climax Molybdenum Co., 1974
4. M.M. Khruschov, The Principles of Abrasive Wear, *Wear,* Vol 28, 1974, p 97-99
5. E. Hornbogen, The Role of Fracture Toughness in the Wear of Metals, *Wear,* Vol 33, 1975, p 251-259
6. K.H. Zum Gahr, Abrasive Wear of Metallic Materials, *Metallurgical Aspects of Wear,* DGM, 1981, p 73-

7. P. Clayton, The Relationship between Wear Behavior and Basic Material Properties for Pearlitic Steels, *Proceedings of the International Conference on Wear of Materials,* American Society of Mechanical Engineers, 1979, p 395-396
8. W.A. Glaeser, High Strain Wear Mechanisms in Ferrous Alloys, *Wear,* Vol 123, 1988, p 155-169
9. N. Jost and I. Schmidt, Friction Induced Martensite in Austenitic Fe-Mn-C Steels, *Proceedings of the International Conference on Wear of Materials,* American Society of Mechanical Engineers, 1985, p 205-214
10. G. Langford, Deformation of Pearlite, *Metall. Trans. A,* 1977, p 861-875
11. A.A. Rosenfield, A Shear Instability Model of Sliding Wear, *Wear,* Vol 116, 1987, p 319-328

Wear of Stainless Steels

John H. Magee, Carpenter Technology Corporation

Introduction

STAINLESS STEELS are primarily used to resist corrosive attack in environments that are as mild as kitchen sinks or as severe as piping used in the chemical process industry. They are broadly defined as steels that contain at least 10.5% Cr.

A wide range of corrosion resistance can be achieved by increasing the chromium content and adding other elements, especially nickel. In addition, high tensile yield strength (>1400 MPa, or 205 ksi) can be accomplished through martensite formation, precipitation hardening, or cold work. Because of the broad definition used, stainless steels are further divided into five families, each with a unique microstructure, alloying element additions, and range of physical, mechanical, and corrosion properties.

The selection of a particular type of stainless steel for an application involves the consideration of such factors as the corrosion resistance of the alloy, mechanical properties, fabricability, and cost. However, for applications such as pumps, valves, bearings, fasteners, and conveyor belts, where one contacting metal surface moves relative to the other, the wear and galling resistance of the metals in contact should also be considered in the selection process.

Stainless steels are characterized as having relatively poor wear and galling resistance, but are often required for a particular application, because of their corrosion resistance. Therefore, finding the most effective alloy to withstand wear and galling can be a difficult problem for design engineers. Lubricants and coatings are often used to reduce wear, although lubricant use is precluded in may applications, such as high-temperature environments, in which they can break down, or food and pharmaceutical processing equipment, which require sanitation.

Additionally, a critical part, such as a valve in a power plant, must resist galling or seizing, because it can shut down or endanger the entire plant (Ref 1, 2).

This article discusses each stainless steel family, specifically in terms of wear resistance. Information on wear and galling, laboratory wear and galling tests, and the associated data from these tests is presented. Applications and design considerations are also discussed.

Classification of Stainless Steels. In the United States, grades of stainless steels are generally designated by one or more of these methods: the American Iron and Steel Institute (AISI) numbering system, the Unified Numbering System (UNS), and proprietary name of designation. In addition, designations have been established by most of the major industrial nations. Of the two institutional numbering systems, AISI is the oldest and most widely used. Most of the grades have a three-digit designation; the 200 and 300 series are generally austenitic stainless steels, whereas the 400

series are either ferritic or martensitic. Some of the grades have a one- or two-letter suffix that indicates a particular modification of the composition.

The UNS system is a broader-based system that comprises a list of metal alloys, including stainless steel. This system includes a considerably greater number of stainless steels than AISI, because it incorporates all of the more recently developed stainless steels. The UNS designation for a stainless consists of the letter S, followed by a five-digit number. For those alloys that have an AISI designation, the first three digits of the UNS designation usually correspond to an AISI number. When the last two digits are 00, the number designates a basic AISI grade. Modifications of the basic grades use two digits other than zeroes.

Table 1 provides the compositional limits for select stainless steels, listed by UNS and AISI type designations and separated into the basic families. Where AISI type designations are not available, common trade names are listed in parentheses. These names, the third commonly used identification of stainless steels, have often become the popular means of identifying a particular alloy. A complete listing of all stainless alloys available is provided in Ref 3.

Table 1 Composition of selected standard and special stainless steels

UNS designation	AISI type	Composition, wt% max									
		C	**Mn**	**Si**	**P**	**S**	**Cr**	**Ni**	**Mo**	**N**	**Others**
Ferritic alloys											
S40500	405	0.08	1.00	1.00	0.040	0.030	11.50-14.50	. . .	. . .	. . .	0.10-0.30 Al
S40900	409	0.08	1.00	1.00	0.045	0.045	10.50-11.75	0.50	. . .	. . .	6 × C-0.75 Ti
S43000	430	0.12	1.00	1.00	0.040	0.030	16.00-18.00	. . .	. . .	. . .	. . .
S43020	430F	0.12	1.25	1.00	0.060	0.1[a]	16.00-18.00	. . .	0.60	. . .	. . .
S43023	430FSe	0.12	1.25	1.00	0.060	0.060	16.00-18.00	. . .	. . .	. . .	0.15 min Se
S43400	434	0.12	1.00	1.00	0.040	0.030	16.00-18.00	. . .	0.75-1.25	. . .	. . .
S44200	442	0.20	1.00	1.00	0.040	0.030	18.00-23.00	. . .	. . .	. . .	. . .
S44300	443[b]	0.20	1.00	1.00	0.040	0.030	18.00-23.00	0.50	. . .	. . .	0.90-1.25 Cu
S44400	444[b]	0.025	1.00	1.00	0.040	0.030	17.50-19.50	1.00	1.75-2.50	0.025	[0.20 + 4 (C + N)]-0.80 Ti + Nb
S44600	446[b]	0.20	1.50	1.00	0.040	0.030	23.00-27.00	. . .	. . .	0.25	. . .
S18200	18-2FM[c]	0.08	1.25-2.50	1.00	0.040	0.15[a]	17.50-19.50	. . .	1.50-2.50	. . .	. . .
Martensitic alloys											
S40300	403	0.15	1.00	0.50	0.040	0.030	11.50-13.00	. . .	. . .	. . .	. . .
S41000	410	0.15	1.00	1.00	0.040	0.030	11.50-13.00	. . .	. . .	. . .	. . .
S41400	414	0.15	1.00	1.00	0.040	0.030	11.50-13.00	1.25-2.50	. . .	. . .	. . .
S41600	416	0.15	1.25	1.00	0.060	0.15[a]	12.00-14.00	. . .	0.60	. . .	. . .
S41610	416 Plus X[d]	0.15	1.50-2.50	1.00	0.060	0.15[a]	12.00-14.00	. . .	0.60	. . .	. . .
S41623	416Se	0.15	1.25	1.00	0.060	0.060	12.00-14.00	. . .	. . .	. . .	0.15 min Se
S42000	420	0.15[a]	1.00	1.00	0.040	0.030	12.00-14.00	. . .	. . .	. . .	. . .
S42010	TrimRite[e]	0.15-0.30	1.00	1.00	0.040	0.030	13.50-15.00	0.25-1.00	0.40-1.00	. . .	. . .
S42020	420F	0.15[a]	1.25	1.00	0.60	0.15[a]	12.00-14.00	. . .	0.60	. . .	. . .

S42023	420FSe[b]	0.30-0.40	1.25	1.00	0.060	0.060	12.00-14.00	. . .	0.60	. . .	0.15 min Se; 0.60 Zr or Cu
S43100	431	0.20	1.00	1.00	0.040	0.030	15.00-17.00	1.25-2.50	. . .	. . .	. . .
S44002	440A	0.60-0.75	1.00	1.00	0.040	0.030	16.00-18.00	. . .	0.75	. . .	. . .
S44003	440B	0.75-0.95	1.00	1.00	0.040	0.030	16.00-18.00	. . .	0.75	. . .	. . .
S44004	440C	0.95-1.20	1.00	1.00	0.040	0.030	16.00-18.00	. . .	0.75	. . .	. . .
S44020	440F[b]	0.95-1.20	1.25	1.00	0.040	0.10-0.35	16.00-18.00	0.75	0.40-0.60	0.08	. . .
S44023	440FSe[b]	0.95-1.20	1.25	1.00	0.040	0.030	16.00-18.00	0.75	0.60	0.08	0.15 min Se
Austenitic alloys											
S20100	201	0.15	5.50-7.50	1.00	0.060	0.030	16.00-18.00	3.50-5.50	. . .	0.25	. . .
S20161	Gall-Tough[e]	0.15	4.00-6.00	3.00-4.00	0.040	0.040	15.00-18.00	4.00-6.00	. . .	0.08-0.20	. . .
S20300	203EZ[f]	0.08	5.00-6.50	1.00	0.040	0.18-0.35	16.00-18.00	5.00-6.50	0.50	. . .	1.75-2.25 Cu
S20910	22-13-5[c]	0.06	4.00-6.00	1.00	0.040	0.030	20.50-23.50	11.50-13.50	1.50-3.00	0.20-0.40	0.10-0.30 Nb; 0.10-0.30 V
S21000	SCF19[e]	0.10	4.00-7.00	0.60	0.030	0.030	18.00-23.00	16.00-20.00	4.00-6.00	0.15	2.00 Cu
S21300	15-15LC[e]	0.25	15.00-18.00	1.00	0.050	0.050	16.00-21.00	3.00	0.50-3.00	0.20-0.80	0.50-2.00 Cu
S21800	Nitronic 60[g]	0.10	7.00-9.00	3.50-4.50	0.040	0.030	16.00-18.00	7.00-9.00	. . .	0.08-0.20	. . .
S21904	21-6-9LC[c]	0.04	8.00-10.00	1.00	0.060	0.030	19.00-21.50	5.50-7.50	. . .	0.15-0.40	. . .
S24100	18-2Mn[c]	0.15	11.00-14.00	1.00	0.060	0.030	16.50-19.50	0.50-2.50	. . .	0.20-0.45	. . .
S28200	18-18 Plus[e]	0.15	17.00-19.00	1.00	0.045	0.030	17.00-19.00	. . .	0.50-1.50	0.40-0.60	0.50-1.50 Cu
. . .	Nitronic 30[g]	0.10	7.00-9.00	1.00	. . .	. . .	15.00-17.00	1.50-3.00	. . .	0.15-0.30	1.00 Cu
S30100	301	0.15	2.00	1.00	0.045	0.030	16.00-18.00	6.00-8.00	. . .	. . .	. . .
S30200	302	0.15	2.00	1.00	0.045	0.030	17.00-19.00	8.00-10.00	. . .	. . .	. . .
S30300	303	0.15	2.00	1.00	0.20	0.15[a]	17.00-19.00	8.00-10.00	0.60	. . .	. . .
S30310	303 Plus X[d]	0.15	2.50-4.50	1.00	0.20	0.25[a]	17.00-19.00	7.00-10.00	0.75	. . .	. . .
S30323	303Se	0.15	2.00	1.00	0.20	0.060	17.00-19.00	8.00-10.00	. . .	. . .	0.15 min Se
S30330	303 Cu[b]	0.15	2.00	1.00	0.15	0.10[a]	17.00-19.00	6.00-10.00	. . .	. . .	2.50-4.00 Cu; 0.10 Se
S30400	304	0.08	2.00	1.00	0.045	0.030	18.00-20.00	8.00-10.50	. . .	. . .	. . .
S30403	304L	0.03	2.00	1.00	0.045	0.030	18.00-20.00	8.00-12.00	. . .	. . .	. . .
S30430	302 HQ[b]	0.10	2.00	1.00	0.045	0.030	17.00-19.00	8.00-10.00	. . .	. . .	3.00-4.00 Cu
S30431	302 HQ-FM[e]	0.06	2.00	1.00	0.040	0.14	16.00-19.00	9.00-11.00	. . .	. . .	1.30-2.40 Cu
S30452	304 HN[b]	0.08	2.00	1.00	0.045	0.030	18.00-20.00	8.00-10.50	. . .	0.16-0.30	. . .
S30500	305	0.12	2.00	1.00	0.045	0.030	17.00-19.00	10.00-13.00	. . .	. . .	. . .
S30900	309	0.20	2.00	1.00	0.045	0.030	22.00-24.00	12.00-15.00	. . .	. . .	. . .
S30908	309S	0.08	2.00	1.00	0.045	0.030	22.00-24.00	12.00-15.00	. . .	. . .	. . .
S31000	310	0.25	2.00	1.50	0.045	0.030	24.00-	19.00-	. . .	. . .	. . .

							26.00	22.00			
S31008	310S	0.08	2.00	1.50	0.045	0.030	24.00-26.00	19.00-22.00	. . .	. . .	. . .
S31600	316	0.08	2.00	1.00	0.045	0.030	16.00-18.00	10:00-4.00	2.00-3.00	. . .	. . .
S31603	316L	0.030	2.00	1.00	0.045	0.030	16.00-18.00	10.00-14.00	2.00-3.00	. . .	. . .
S31620	316F	0.08	2.00	1.00	0.20	0.10[a]	17.00-19.00	12.00-14.00	1.75-2.50	. . .	. . .
S31700	317	0.08	2.00	1.00	0.045	0.30	18.00-20.00	11.00-15.00	3.00-4.00	. . .	. . .
S31703	317L	0.030	2.00	1.00	0.045	0.030	18.00-20.00	11.00-15.00	3.00-4.00	. . .	. . .
S32100	321	0.08	2.00	1.00	0.045	0.030	17.00-19.00	9.00-12.00	. . .	. . .	5 × C min Ti
S34700	347	0.08	2.00	1.00	0.045	0.030	17.00-19.00	9.00-13.00	. . .	. . .	10 × C min Nb
S34720	347F[b]	0.08	2.00	1.00	0.045	0.18-0.35	17.00-19.00	9.00-12.00	. . .	. . .	10 × C-1.10 Nb
S34723	347FSe[b]	0.08	2.00	1.00	0.11-0.17	0.030	17.00-19.00	9.00-12.00	. . .	. . .	10 × C-1.10 Nb; 0.15-0.35 Se
S38400	384	0.08	2.00	1.00	0.045	0.030	15.00-17.00	17.00-19.00	. . .	. . .	. . .
N08020	20Cb-3[e]	0.07	2.00	1.00	0.045	0.035	19.00-21.00	32.00-38.00	2.00-3.00	. . .	8 × C-100 Nb; 3.00-4.00 Cu
Duplex alloys											
S31803	2205[c]	0.30	2.00	1.00	0.030	0.020	21.0-23.0	4.50-6.50	2.50-6.50	0.08-0.20	. . .
S32550	Alloy 255[c]	0.04	1.50	1.00	0.04	0.03	240-27.0	4.50-6.50	2.00-4.00	0.10-0.25	1.50-2.50 Cu
S32900	329	0.20	1.00	0.75	0.040	0.030	23.00-28.00	2.50-5.00	1.00-2.00	. . .	. . .
S32950	7-Mo Plus[e]	0.03	2.00	0.60	0.035	0.010	26.0-29.0	3.50-5.20	1.00-2.50	0.15-0.35	. . .
Precipitation-hardenable alloys											
S13800	PH13-8 Mo[g]	0.05	0.20	0.10	0.010	0.008	12.25-13.25	7.50-8.50	2.00-2.50	0.01	0.90-1.35 Al
S15500	15-5PH[g]	0.07	1.00	1.00	0.040	0.030	14.00-15.50	3.50-5.50	. . .	. . .	0.15-0.45 Nb; 2.50-4.50 Cu
S15700	15-7PH[g]	0.09	1.00	1.00	0.040	0.030	14.00-16.00	6.50-7.25	2.00-3.00	. . .	0.75-1.50 Al
S17400	17-4PH[g]	0.07	1.00	1.00	0.040	0.030	15.50-17.50	3.00-5.00	. . .	. . .	0.15-0.45 Nb; 3.00-5.00 Cu
S17700	PH 17-7[g]	0.09	1.00	1.00	0.040	0.040	16.00-18.00	6.5-7.75	. . .	. . .	0.75-1.50 Al
S35000	633[b]	0.07-0.11	0.50-1.25	0.50	0.040	0.030	16.00-17.00	4.00-5.00	2.50-3.25	0.07-0.13	. . .
S35500	634[b]	0.10-0.15	0.50-1.25	0.50	0.040	0.030	15.00-16.00	4.00-5.00	2.50-3.25	0.07-0.13	. . .
S45000	Custom 450[e]	0.05	1.00	1.00	0.030	0.030	14.00-16.00	5.00-7.00	0.50-1.00	. . .	8 × C min; 1.25-1.75 Cu
S45500	Custom 455[e]	0.05	0.50	0.50	0.040	0.030	11.00-12.50	7.50-9.50	0.50	. . .	0.10-0.50 Nb; 1.50-2.50 Cu 0.80-1.40 Ti
S66286	A286[c]	0.08	2.00	1.00	0.040	0.030	13.50-16.00	24.0-27.0	1.00-1.50	. . .	0.35 Al; 0.0010-0.010 B 1.90-2.35 Ti; 0.10-0.50 V

Note: All compositions include Fe as balance.

(a) Minimum, rather than maximum wt%.

(b) Designation resembles AISI type, but is not used in that system.

(c) Common trade name, rather than AISI type.

(d) Trade name of Crucible Inc.

(e) Trade name of Carpenter Technology Corporation.

(f) Trade name of Al-Tech Corp.

(g) Trade name of Armco Inc.

Families of Stainless Steels

On the basis of microstructure, there are five major families of stainless steels: ferritic, austenitic, martensitic, precipitation-hardenable (PH), and duplex (austenite and ferrite).

Ferritic stainless steels are so named because their body-centered-cubic (bcc) crystal structure is the same as iron at room temperature. These alloys are magnetic and cannot be hardened by heat treatment. In general, ferritic stainless steels do not have particularly high strength. Their annealed yield strengths range from 275 to 350 MPa (40 to 50 ksi), and their poor toughness and susceptibility to sensitization limit their fabricability and the usable section size. Their chief advantages are their resistance to chloride stress-corrosion cracking, atmospheric corrosion, and oxidation, at a relatively low cost.

Ferritic stainless steels contain between 11 and 30% Cr, with only small amounts of austenite-forming elements, such as carbon, nitrogen, and nickel. Their general use depends on their chromium content.

The low-chromium (11%) alloys (S40500 and S40900) have fair corrosion and oxidation resistance and good fabricability at low cost. They have gained wide acceptance for use in automotive exhaust systems.

The intermediate-chromium (16 to 18%) alloys (S43000 and S43400) are used for automotive trim and cooking utensils. These alloys are not as readily fabricated as the lower Cr alloys, because of their poor toughness and weldability.

The high-chromium (19 to 30%) alloys (S44200 and S44600) are used for applications that require a high level of corrosion and oxidation resistance. These alloys often contain either aluminum or molybdenum and have a very low carbon content. Their fabrication is possible because of special melting techniques that can achieve very low carbon, as well as very low nitrogen contents. Stabilizing elements, like titanium and niobium, can be added to prevent sensitization and to improve as-welded properties.

Austenitic stainless steels constitute the largest stainless family, in terms of number of alloys and usage. Like the ferritic alloys, they cannot be hardened by heat treatment. However, their similarity ends there. The austenitic alloys are nonmagnetic, and their structure is face-centered-cubic (fcc), like high-temperature (900 to 1400 °C, or 1650 to 2550 °F) iron. They possess excellent ductility, formability, and toughness, even at cryogenic temperatures. In addition, they can be substantially hardened by cold work.

Although nickel is the chief element used to stabilize austenite, carbon and nitrogen are also used, because they are readily soluble in the fcc structure. A wide range of corrosion resistance can be achieved by balancing the ferrite-forming elements, such as chromium and molybdenum, with austenite-forming elements.

Austenitic stainless steel can be subdivided into two categories: chromium-nickel alloys, such as S30400 and S31600, and chromium-manganese-nitrogen alloys, such as S20100 and S24100. The latter group generally contains less nickel and maintains the austenitic structure with high levels of nitrogen. Manganese (5 to 20%) is necessary in these low-nickel alloys to increase nitrogen solubility in austenite and to prevent martensite transformation. The addition of nitrogen also increases the strength in austenitic alloys. Typical chromium-nickel alloys have tensile yield strengths from 200 to 275 MPa (30 to 40 ksi) in the annealed condition, whereas the high-nitrogen alloys have yield strengths up to 500 MPa (70 ksi).

As previously mentioned, austenitic alloys can be substantially hardened by cold working. The degree of work hardening depends on alloy content, with increasing alloy content decreasing the work-hardening rate. Figure 1 depicts the higher work-hardening rate of type 301 (7% Ni) versus type 305 (11.5% Ni), which is primarily due to its lower nickel content. Austenitic stainless steels that have a low alloy content, such as S20100, S20161, S30100, and S30400, often become magnetic because of the transformation to martensite when sufficiently cold worked or heavily deformed in machining or

forming operations. The rapid work hardening of S20161 is a major advantage in sliding wear. In S30430, copper is intentionally added to lower the work-hardening rate for enhanced headability in the production of fasteners.

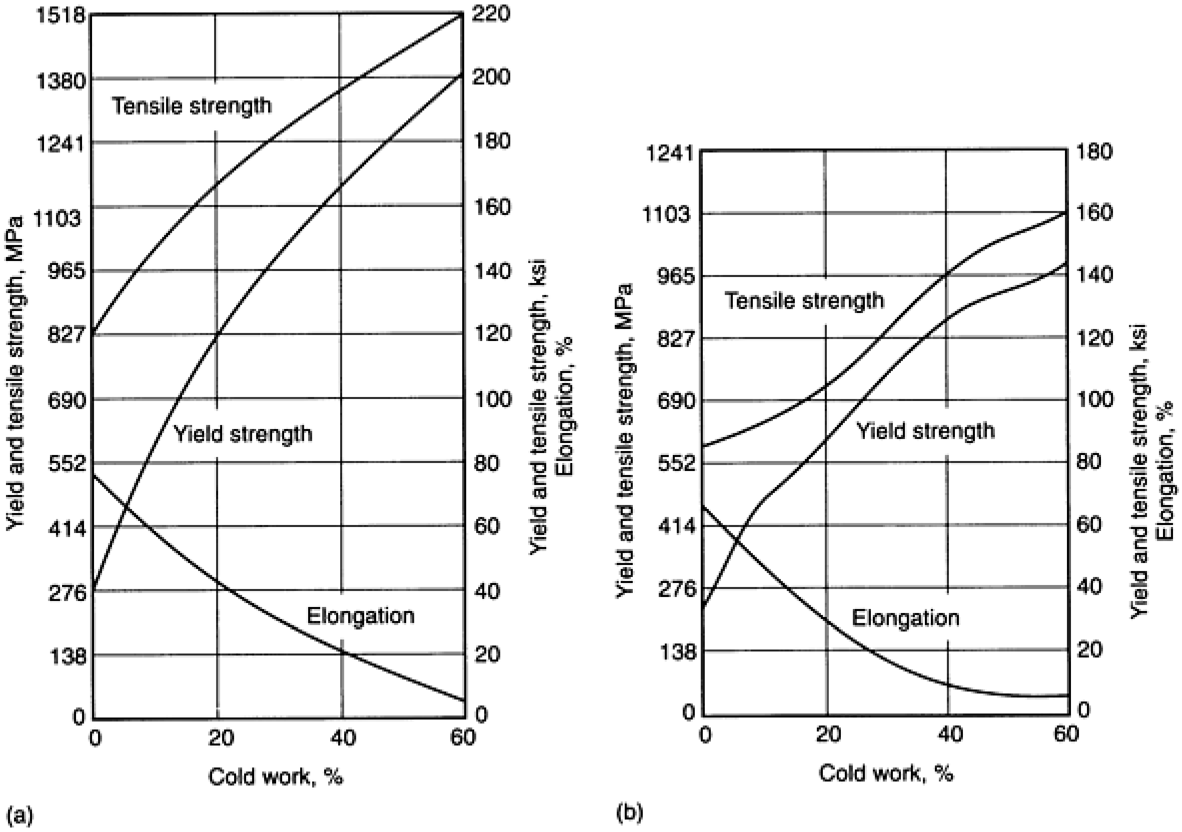


Fig. 1 Effect of cold working on mechanical properties of stainless steels. (a) Type 301. (b) Type 305. Source: Ref 4

Another property that depends on alloy content is corrosion resistance. Molybdenum is added to S31700 and S31600 to enhance corrosion resistance in chloride environments. High-chromium grades (S30900 and S31000) are used in oxidizing environments and high-temperature applications, whereas a high-nickel grade (N08020) is used in severe reducing acid environments. To prevent intergranular corrosion after elevated-temperature exposure, titanium or niobium is added to stabilized carbon in S32100 or S34700. Also, lower-carbon grades (AISI L or S designations), such as S30403 (type 304L), have been established to prevent intergranular corrosion.

Martensitic stainless steels are similar to iron-carbon alloys that are austenitized, hardened by quenching, and then tempered for increased ductility and toughness. These alloys are magnetic, and their heat-treated structure is body-centered tetragonal. In the annealed condition, they have a tensile yield strength of about 275 MPa (40 ksi) and are generally machined, cold formed, and cold worked in this condition.

The strength obtained by heat treatment depends on the carbon content of the alloy. Increasing carbon content increases strength, but decreases ductility and toughness. The most commonly used alloy in this family is S41000, which contains about 12% Cr and 0.1% C. This alloy is tempered to a variety of hardness levels, from 20 to 40 HRC. Both chromium and carbon contents are increased in alloys S42000, S44002, S44003, and S44004. The first of these contains 14% Cr and 0.3% C and has a hardness capability of 50 HRC. The other three alloys contain 16% Cr and from 0.6 to 1.1% C. These alloys are capable of 60 HRC, and a tensile yield strength of 1900 MPa (280 ksi). The amount of primary carbides increases with increased carbon content in these three alloys.

Wear resistance for martensitic stainless steels is very dependent on carbon content. S44004 (1.1% C) has excellent adhesive and abrasive wear, similar to tool steels, whereas S41000 (0.1% C) has relatively poor wear resistance. The key

to adhesive wear resistance is a high hardness. Abrasive wear resistance requires both high hardness and primary carbides.

Molybdenum and nickel can be added to martensitic stainless steel to improve corrosion and toughness properties. However, the addition of these elements is somewhat restricted, because higher amounts results in a microstructure that is not fully martensitic.

PH stainless steels are chromium-nickel grades that can be hardened by an aging treatment. These grades are classified as austenitic (such as S66286), semi-austenitic (such as S17700), or martensitic (such as S17400). The classification is determined by their solution-annealed microstructure. The semi-austenitic alloys are subsequently heat treated, so that the austenite transforms to martensite. Cold work is sometimes used to facilitate the aging reaction. Various alloying elements, such as aluminum, titanium, niobium, or copper, are used to achieve aging. They generally form intermetallic compounds, but in S17400, fine copper precipitates are formed.

Like the martensitic stainless steels, PH alloys can attain high tensile yield strengths, up to 1700 MPa (250 ksi). However, these alloys have superior ductility, toughness, and corrosion resistance, compared with the martensitic alloys. These properties are related to their higher chromium, nickel, and molybdenum contents, as well as their restricted carbon (0.040 max.) levels. The low carbon content of the martensitic PH stainless steels is especially critical for toughness and good ductility. However, this low carbon content reduces the wear resistance of these alloys.

Duplex stainless steels are chromium-nickel-molybdenum alloys that are balanced to contain a mixture of austenite and ferrite, and are magnetic, as well. Their duplex structure results in improved stress-corrosion cracking resistance, compared with the austenitic stainless steels, and improved toughness and ductility, compared with the ferritic stainless steels. They are capable of tensile yield strengths ranging from 550 to 690 MPa (80 to 100 ksi) in the annealed condition, which is approximately twice the strength level of either phase alone.

The original alloy in this family was the predominantly ferritic S32900. The addition of nitrogen to duplex alloys, such as S32950 and S31803, increases the amount of austenite to nearly 50%. In addition, nitrogen improves aswelded corrosion properties, chloride corrosion resistance, and toughness. The improvement in toughness is probably related to the higher amount of austenite present, which makes it possible to produce heavier product forms, such as plates and bars.

Physical and Mechanical Properties of Stainless Steels

The physical and mechanical properties of stainless steels are quite different from those of aluminum and copper alloys. However, when comparing the various stainless families with carbon steels, many similarities in properties exist, although there are some key differences. Like carbon steels, the density of stainless steels is $\sim$8.0 g/cm^3, which is approximately three times greater than that of aluminum alloys (2.7 g/cm^3). Like carbon steels, stainless steels also have a high modulus of elasticity (200 MPa, or 30 ksi), which is nearly twice that of copper alloys (115 MPa, or 17 ksi) and nearly three times that of aluminum alloys (70 MPa, or 10 ksi).

Differences between these materials are evident in thermal conductivity, thermal expansion, and electrical resistivity, as well. Figure 2 shows the large variation in thermal conductivity between various types of materials: type 6061 aluminum has a very high thermal conductivity, followed by aluminum bronze, 1080 carbon steel, and then stainless steels. For stainless steels, alloying additions, especially nickel, copper, and chromium, greatly decrease thermal conductivity.

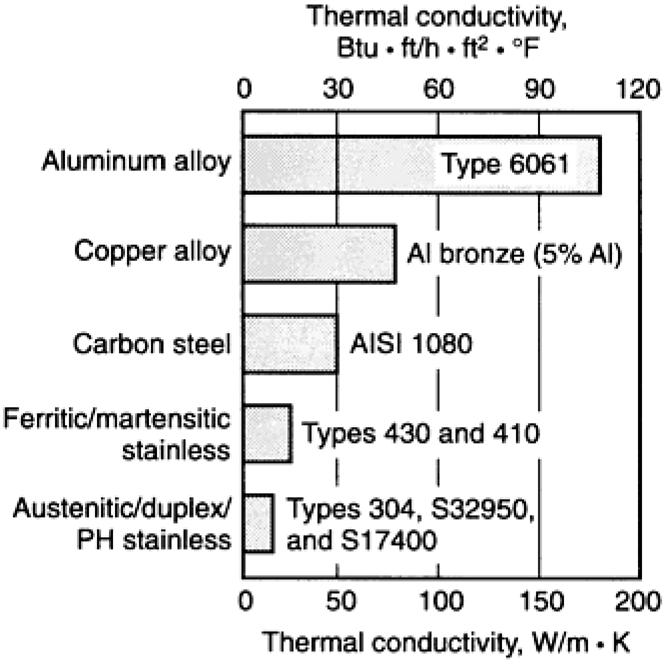


Fig. 2 Comparison of thermal conductivity for carbon steel, copper alloy, aluminum, and stainless steels. Source: Ref 5, 6

Thermal expansion (Fig. 3) is greatest for type 6061 aluminum alloy, followed by aluminum bronze and austenitic stainless alloys, and then ferritic and martensitic alloys. For austenitic stainless alloys, additions of Ni and Cu can decrease thermal expansion. Stainless steels have high electrical resistivity (Fig. 4). Alloying additions tend to increase electrical resistivity. Therefore, the ferritic and martensitic stainless steels have lower electrical resistivity than the austenitic, duplex, and PH alloys, but higher electrical resistivity than 1080 carbon steel. Electrical resistivity of stainless steels is ~7.5 times greater than aluminum bronze and nearly 20 times greater than type 6061 aluminum alloy (Ref 5, 6).

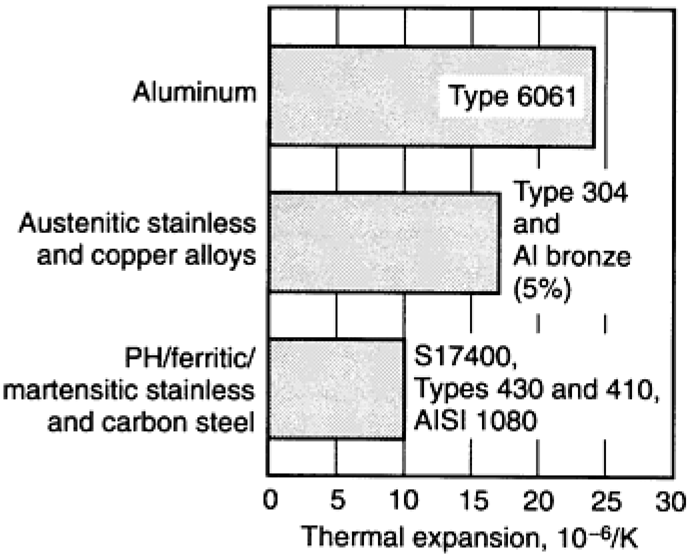


Fig. 3 Comparison of thermal expansion for carbon steel, copper alloy, aluminum, and stainless steels. Source: Ref 5, 6

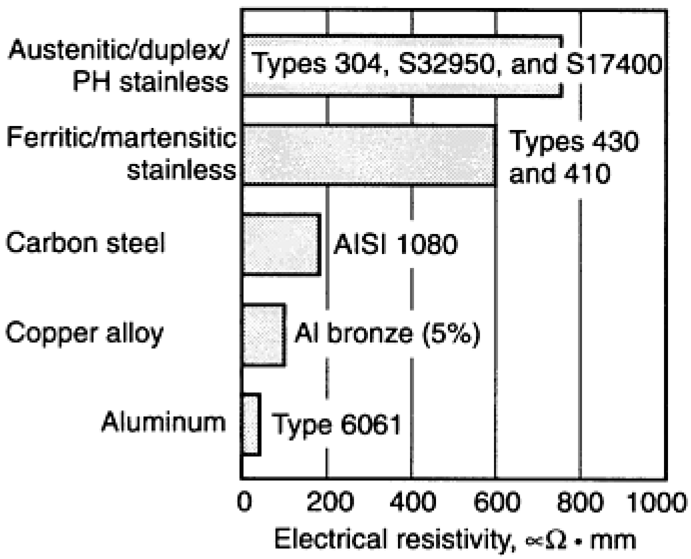


Fig. 4 Comparison of electrical resistivity for carbon steel, copper alloy, aluminum, and stainless steels. Source: Ref 5, 6

Table 2 lists tensile properties and toughness for selected stainless alloys representing the five families. The four grades listed under austenitic alloys have relatively low yield strength, compared with the heat-treatable alloys, but have the highest tensile ductility and toughness. The latter two alloys, S20161 and S21800, were specifically developed to have superior resistance to galling and metal-to-metal wear for stainless steels. Alloy N08020 is a high-nickel (33%) stainless alloy for use in harsh corrosive environments.

Table 2 Properties of selected stainless steels relative to various ferrous and nonferrous alloys

UNS or AISI type	Condition	Rockwell hardness	Average tensile properties						Charpy V-notch impact strength	
			Yield strength, 0.2% offset		**Ultimate tensile strength**		**Elongation in 50.8 mm (2.0 in.), %**	**Reduction of area, %**		
			MPa	**ksi**	**MPa**	**ksi**			**J**	**ft · lbf**
Austenitic stainless										
Type 304	Annealed	81 HRB	241	35	586	85	60.0	70.0	≥325	≥240
N08020	Annealed	84 HRB	276	40	621	90	50.0	65.0	271	200
S20161	Annealed	93 HRB	365	53	970	140	59.0	64.0	≥325	≥240
S21800	Annealed	95 HRB	414	60	710	103	64.0	74.0	≥325	≥240
Ferritic										
Type 405	Annealed	81 HRB	276	40	483	70	30.0	60.0	. . .	. . .
Type 430	Annealed	82 HRB	310	45	517	75	30.0	65.0	217	161
Duplex										
S32950	Annealed	100 HRB	570	82	760	110	38.0	78.0	157	116
Martensitic										
Type 410	Annealed	82 HRB	276	40	517	75	35.0	70.0	. . .	. . .
	Oil quenched from 1010 °C (1850 °F) and tempered:									
	at 250 °C (500 °F)	43 HRC	1089	158	1337	193	17.0	62.0	76	56
	at 593 °C (1100 °F)	26 HRC	724	105	827	120	20.0	63.0	52	38
Type 420	Annealed	92 HRB	345	50	655	95	25.0	55.0	. . .	. . .
	Oil quenched from 1038 °C (1900 °F) and tempered at 316 °C (600 °F)	52 HRC	1482	215	1724	250	8.0	25.0	20	15
Type 440C	Annealed	97 HRB	448	65	758	110	14.0	30.0	. . .	. . .
	Oil quenched from 1038 °C (1900 °F) and tempered at 316 °C (600 °F)	57 HRC	1896	275	1975	285	2.0	10.0	2	3
Precipitation hardened										
S45500	Annealed	31 HRC	793	115	1000	145	14.0	70.0	. . .	. . .
	Water quenched from 1038 °C (1900 °F) and aged:									
	at 482 °C (900 °F)	49 HRC	1620	235	1689	245	10.0	45.0	12	9
	at 566 °C (1050 °F)	40 HRC	1207	175	1310	190	15.0	55.0	47	35
S17400	Annealed	31 HRC	793	115	965	140	12.0	50.0	. . .	. . .
	Water quenched from 1038 °C (1900 °F) and aged:									
	at 482 °C (900 °F)	44 HRC	1262	183	1365	198	15.0	52.0	21	16
	at 621 °C (1150 °F)	33 HRC	869	126	1131	164	17.0	59.0	75	55
Carbon steel										
AISI 1080	Annealed	97 HRB	455	66	821	119	15.0	22.0	. . .	. . .
	Oil quenched from 816 °C (1500 °F) and tempered at 204 °C (400 °F)	42 HRC	980	142	1304	189	12.0	35.0	. . .	. . .
Aluminum alloy										
Type 6061	Annealed	. . .	55	8	124	18	25.0	. . .	. . .	. . .
	Aged	56 HRB	276	40	311	45	12.0	. . .	. . .	. . .
Copper alloy										
Al bronze (95Cu-5Al)	Annealed	45 HRB	173	25	380	55	65.0	. . .	. . .	. . .

Source: Ref 5, 6

The ferritic stainless steels (type 405 and 409) listed have tensile yield strengths similar to the austenitic grades, but lower values for ultimate tensile strength, ductility, and toughness. However, strength, ductility, and toughness are still excellent, compared with other materials, such as type 6061 aluminum and aluminum bronze. The duplex stainless alloy (S32950) listed has twice the tensile yield strength of the austenitic and ferritic grades, and approximately half the toughness. Again, its toughness is far superior to heat-treat-hardened alloys.

The martensitic alloys listed in Table 2 have a large variation in strength, ductility, and toughness. In the annealed condition, their properties are similar to the ferritic alloys, with strength increasing and ductility decreasing with increasing carbon content. The higher carbon-containing alloys, type 420 and type 440C, are generally tempered at a low temperature (330 °C, or 625 °F max.) to maximize their strength. On the other hand, type 410 is tempered over a wide temperature range, from 260 to 650 °C (500 to 1200 °F). The tensile properties of type 410 are similar to carbon steel (AISI 1080).

The martensitic PH alloys, such as S45500 and S17400, have higher annealed strength and lower ductility than the martensitic alloys, and are aged at temperatures ranging from 480 to 620 °C (895 to 1150 °F). Their strength is dependent on the hardener (Ti, Nb, Cu), the amount of hardener, and the aging temperatures used. Toughness is either similar to superior to the martensitic alloys at a given strength level.

Wear and Galling

Types of Wear

The types of wear described below include abrasive, fretting, corrosive, fatigue, and adhesive wear.

Abrasive wear involves the plowing of localized surface contacts through a softer mated material. The wear is most frequently caused by nonmetallic materials, but metallic particles can also cause abrasion (Ref 7). Generally, a material is seriously abraded or scratched only by a particle harder than itself. Abrasive wear is commonly divided into three types: low stress, high stress, and gouging.

Low-stress abrasion (scratching) is defined as wear that occurs due to relatively light rubbing contact of abrasive particles with the metal. Wear scars usually show scratches, and the amount of subsurface deformation is minimal. Consequently, the surface does not work harden appreciably. Parts such as screens, chute liners, blades, and belts that are exposed to sand slurries or abrasive atmospheres could experience low-stress abrasion. Many machine components such as bushings, seals, and chains that operate in dust will wear by low-stress abrasion (Ref 8).

High-stress abrasion is wear under a level of stress that is high enough to crush the abrasive. Considerably more strain hardening of the metal surface occurs. The abrasion or ore grinding balls is an example of high-stress abrasion in the mining industry. Other examples include abrasion experienced by rolling-contact bearings, gears, cams, and pivots.

Gouging Abrasion. The term gouging abrasion is used to describe high-stress abrasion that results in sizable grooves or gouges on the worn surface (Ref 9). It occurs on parts such as crusher liners, impact hammers in pulverizers, and dipper teeth handling large rocks. Strain hardening and deformation are the dominant factors.

For ferrous materials, abrasion resistance is highly dependent on three metallurgical variables: microstructure, hardness, and carbon content. The inherently hard martensitic structure is preferable to the softer ferritic and austenitic structures. This is especially significant in low-stress abrasion, where little subsurface deformation occurs. When high-stress abrasion is encountered, alloys with high work-hardened hardness values have improved wear resistance, when compared with alloys with low work-hardened hardness values. Although austenitic stainless less steels will work harden more readily than the other stainless families of alloys, martensitic stainless alloys are preferred in applications where high-stress abrasion is encountered, because of their higher hardness by heat treatment. Increased carbon content, regardless of structure, favors better abrasive wear resistance, and so does an increased volume of carbides, as long as their hardness is not exceeded by that of the abrasive medium.

For stainless steels, knowledge of low-stress abrasion resistance is important, because abrasive particles can be found in applications where stainless steels are used. Austenitic stainless steels with a high work-hardening rate could be used where gouging abrasion is a problem and toughness is required. However, austenitic manganese steels, such as Hadfield Mn steel, are more resistant. Generally, stainless steels are considered for abrasive wear conditions when the environment is corrosive or when elevated temperature are encountered.

Fretting wear is material loss that is due to very small amplitude vibrations at mechanical connections, such as riveted joints. This type of wear is a combination of oxidation and abrasive wear. Oscillation of two metallic surfaces produces tiny metallic fragments that oxidize and become abrasive. Subsequent wear proceeds by mild adhesive wear in combination with abrasive wear.

Fretting wear is influenced by contact conditions, environmental conditions, and material properties and behavior. These factors may interact to influence both the nature and the extent of fretting damage. For example, the influence of an environmental factor depends on its accessibility to the metal contact area. Only if the environmental conditions have ready access to fretting damage sites, then environmental factors will strongly influence fretting.

Key parameters in fretting include load, frequency, amplitude of fretting motion, number of fretting cycles, relative humidity, and temperature. Fretting wear rate is virtually independent of amplitude up to a critical value. Beyond that, the wear rate increases almost linearly with the amplitude. The effect of frequency has been studied on mild steel. Up to 30 Hz, the fretting wear decreases with increasing frequency, while wear is not affected above 30 Hz. A threshold number of cycles appears to be required for the onset of steady-state fretting wear rate. This is marked by the appearance of microspall pits, which indicate that a surface fatigue mechanism is operative. The environmental factors, that is, relative humidity and temperature, generally favor the use of stainless steels, because of their superior corrosion and thermal properties.

Corrosive wear involves an interaction between the wear surface and the corrosive reagent. Corrosion in aqueous media is an electrochemical action that results in material removal by dissolution, whereas wear involves material removal that is due to physical interaction between surfaces under relative motion. When these two processes are combined, the material loss may be significantly accelerated, because of synergistic behavior. The wear-corrosion process involves the disruption and removal of the oxide film, leading to exposure of active metal surface to the environment, dissolution or repassivation of the exposed metal surface, interaction between elastic fields at asperities in contact with the environment, and interaction between plastically deformed areas and the environment. The exact nature of the wear-corrosion process is very dependent on the specific metal/corrosive reagent (Ref 10).

Corrosive wear can occur in the mining industry, where abrasive wear combines with a wet corrosive environment. Abrasion-resistant alloy steels can be ineffective in this application, whereas stainless steels often perform well. In these conditions, carbon steel readily forms iron oxide, which is removed by the sliding and bumping of moving coal/ore (Ref 11). If the rust is repeatedly removed, then there will be continuing loss of metal thickness.

In South Africa, a ferritic grade (3Cr12) provides a cost-effective solution to corrosive wear problems in chutes, liners, and conveyor belt equipment used in ore handling. This alloy is a modified type 409 stainless steel. In the United States, type 304 has been widely used in coal-handling equipment such as chutes, bins, hoppers, and screens, because of its good corrosion resistance. Another reason for the popularity of type 304 is the great improvement over alloy steels with regard to "slideability." Type 304 retains a smooth surface finish, whereas abrasion-resistant alloy steels rust, which causes material buildup and lowers flow rate. Coal hopper cars lined with type 304 can be discharged three times faster than unlined cars, because of improved slideability (Ref 12).

Corrosive wear is clearly a situation in which the use of stainless steels is attractive, because of their ability to resist the removal of their oxide film.

Fatigue wear, or contact fatigue, occurs when a surface is stressed in a cyclic manner. This type of wear can be found in parts subjected to rolling contact, such as ball bearings and gears. The fatigue wear rate of metals is affected by surface conditions, such as finish, residual stress, hardness, and microstructure. Surface treatments such as nitriding, carburizing, and shot peening, which increase surface hardness and improve residual stress distribution, are performed to prevent fatigue wear.

Cavitation, a combination of corrosive wear and fatigue wear, occurs when a liquid is subjected to rapidly alternating changes in pressure, during which bubbles form. Subsequent equalization of pressure causes the bubbles to collapse at the liquid-metal interface, resulting in cavitation wear. High-velocity pumps, hydraulic turbines, and fluid flow valves are applications where cavitation is a concern (Ref 13).

Because of the corrosive aspect of cavitation, austenitic stainless steels that work harden rapidly can be used satisfactorily. A cobalt-base hardfacing alloy, such as Stellite 6B, is used to achieve the best resistance.

Adhesive wear occurs when two metallic components slide against each other under an applied load where no abrasives are present. This type of wear is called "adhesive" because of the strong metallic bonds formed between surface asperities (that is, surface high spots) of the materials. Wear results from the shear failure of the weaker of the two metallic mating surfaces. One theory postulates that subsurface crack nucleation and growth follow asperity shearing and flattening (Ref 14).

When the applied load is low enough, the surface oxide film characteristic of stainless steels can prevent the formation of metallic bonds between the asperities on the sliding surfaces, resulting in low wear rates. This form of wear is called mild wear, or oxidative wear, and can be tolerated by most moving components. When the applied load is high, metallic bonds will form between the surface asperities, and the resulting wear rates are high. The load at which there is a transition from mild to severe wear is called the transition load.

Adhesive wear is more prevalent in parts where a lubricant cannot be used. Examples include chain-link conveyor belts, fasteners, and sliding components in a valve. For stainless steels, hardness affects adhesive wear resistance. For martensitic alloys, a minimum hardness of 53 HRC is required for excellent wear resistance. For austenitic stainless alloys, the work-hardened hardness is critical, as are alloying additions that increase the stability of the oxide film. These factors tend to increase the transition load required for severe wear to occur.

Galling

Galling can be considered a severe form of adhesive wear. With high loads and poor lubrication, surface damage can occur on sliding metal components. The damage is characterized by localized macroscopic material transfer, that is, large fragments or surface protrusions that are easily visible on either or both surfaces. This gross damage is usually referred to as galling, and it can occur after just a few cycles of movement between the mating surfaces. Severe galling can result in seizure of the metal surfaces.

The terms scuffing and scoring are also used to describe similar surface damage under lubricated conditions. Scuffing is the preferred term when the damage occurs at lubricated surfaces, such as the piston ring-cylinder wall contact. Scoring typically describes damage that takes the form of relatively long grooves (Ref 15).

Materials that have limited ductility are less prone to galling, because under high loads surface asperities will tend to fracture when interlocked. Small fragments of material may be lost, but the resultant damage will be more similar to scoring than to galling. For highly ductile materials, asperities tend to plastically deform, thereby increasing the contact area of mated surfaces; eventually, galling occurs.

Another key material behavior during plastic deformation is the ease with which dislocations cross slip over more than one plane. In fcc materials, such as austenitic stainless steels, dislocations easily cross slip. The rate of cross slip for a given alloy or element is usually indicated by its stacking-fault energy. Dislocation cross slip is hindered by the presence of stacking faults, and a high stacking-fault energy indicates a low number of impeding stacking faults and an increased tendency to cross slip and, hence, gall. Table 3 lists the stacking-fault energies of four fcc elements. Nickel and aluminum have poor galling resistance, whereas gold and copper have good galling resistance. Austenitic stainless steels with high work-hardening rates will have relatively low stacking-fault energies, and have been shown to have less tendency to gall.

Table 3 Stacking-fault energies of some common metals

Metals	Stacking-fault energy, eV gs/cm^2
Gold	30
Copper	40
Nickel	80
Aluminum	200

Source: Ref 15, p 183

Materials that have a hexagonal close-packed (hcp) structure with a high *c/a* ratio have a low dislocation cross slip rate and are less prone to galling. This explains why cobalt-base alloys and cadmium-plated alloys resist galling while titanium alloys tend to gall.

Factors Affecting Wear and Galling

The factors that affect wear and galling can be design, lubrication, environmental, and material related. Component design is probably the most critical factor. When stainless steels are required, proper design can minimize galling and wear. Similar applications, like valve parts, can often result in wear-related problems for one company or be of very little concern for another, despite their use of the same alloy.

When stainless steels are used in sliding surface applications, a key consideration is reduction of contact stress. The load on the parts should be minimized, and contact area should be maximized. Design tolerance of the parts should be tight with sufficient clearance because tightly fit parts will be more prone to wear and galling. Lubricants should be used where possible, because they are very effective at reducing contact stress. However, the design should ensure that lubrication can be effectively applied. Often, lubricants are ineffective because of poor design, which renders the parts inacessible to lubrication.

Another important factor is the surface roughness of the parts. Highly polished surfaces (<0.25 μm, or 10 μin.) or very rough finishes (>1.5 μm, or 60 μin.) increase the tendency for wear and galling. It is theorized that very smooth surfaces lack the ability to store wear debris, because of the absence of valleys between asperities, which means the asperities will have greater interaction. Also, lubricants will tend to wipe off the smoother surface. Too rough a finish results in interlocking asperities, which promote severe tearing and galling.

A final design option is surface treatments, such as nitriding, carburizing, hard-face coatings, ion implantation, and shot peening. These treatments are effective at reducing wear and galling, provided that the part configuration or the added cost is acceptable.

Wear and galling can also be affected by the environment. Stainless steels are selected instead of carbon steels when the environment requires either oxidation or corrosion resistance. Alloy selection is dependent on the specific environment that the parts will encounter in service. Thus, laboratory and field tests of candidate alloys are highly recommended. Once again, component design is a critical factor in reducing corrosive wear.

In the mining industry, proper design of equipment can reduce corrosive wear, thereby increasing service life. Sharp bends in flames and pipelines should be avoided, and the angle of discharge from belts or chutes should be minimized. Another consideration is the possibility of galvanic corrosion, which can occur where dissimilar metals are connected in the presence of an electrolyte. An example would be the use of carbon steel fasteners to join stainless components. The dissimilar metal combination, plus the adverse surface area ratio between the fastener and the surrounding surface, will accelerate the corrosion of the carbon steel.

The proper design of equipment should allow for the drainage of pipes, flumes, and tanks during shutdown, as well as spillage or hosedown water within the plant. Examples of proper versus improper design of coal-handling equipment are shown in Fig. 5.

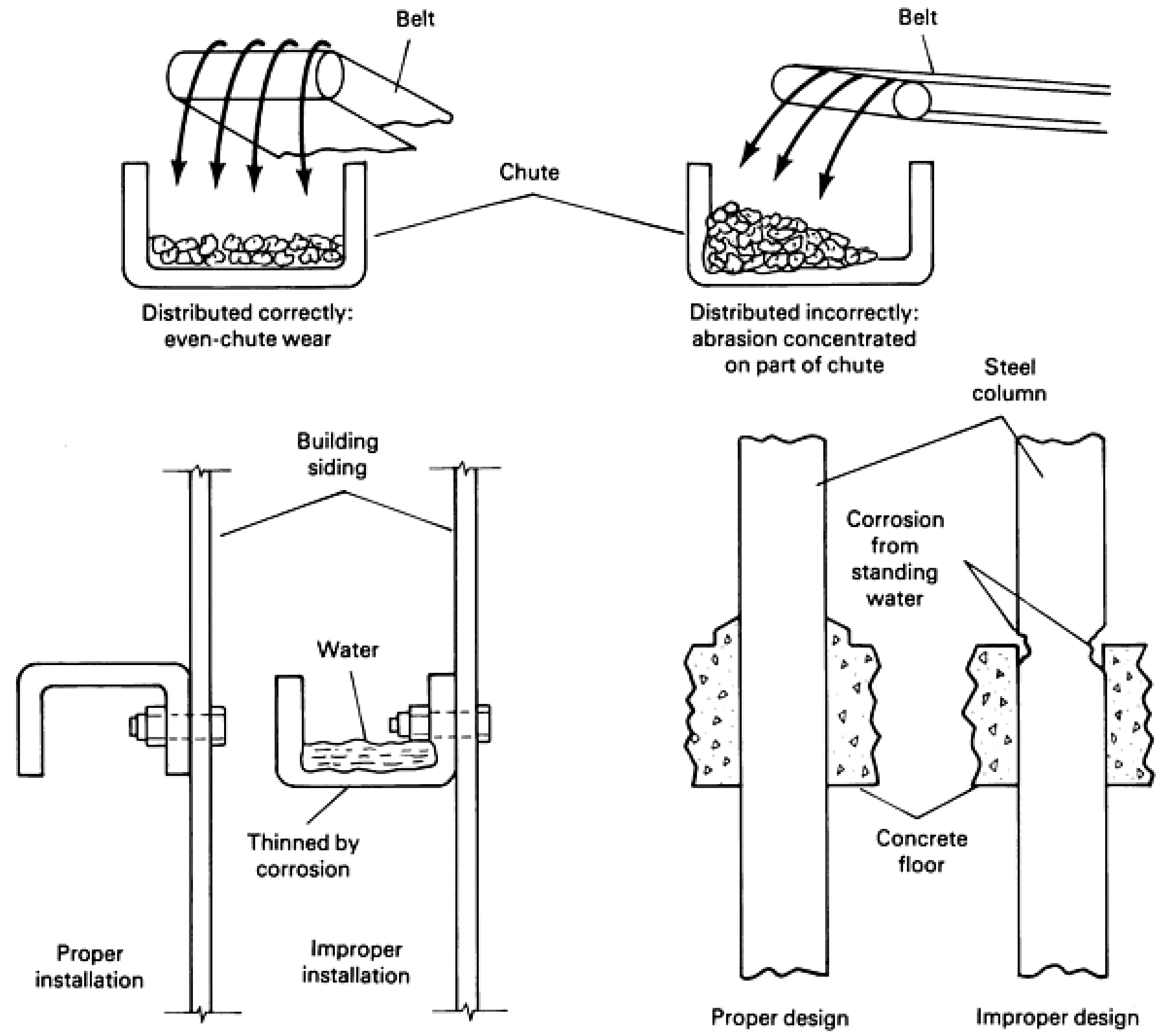


Fig. 5 Effect of design on susceptibility to corrosive wear. Source: Ref 11

Material selection is dependent on the type of wear encountered. To resist adhesive wear, a high work-hardened hardness in an austenitic alloy and a stable oxide film can have excellent results. For hardenable stainless steels, high carbon content (0.3% minimum) to increase hardness (53 HRC minimum) is critical. Thus, martensitic PH alloys with their low carbon content, but high hardness, tend to have poor wear resistance. Because ferritic stainless steels cannot be hardened by heat treatment nor readily work hardened, they may also have poor metal-to-metal wear resistance.

The comments on adhesive wear apply to galling resistance, except that the martensitic PH alloys with high hardness, which is achieved by using a low aging temperature (480 °C, or 895 °F), can have improved galling resistance. Elements that form inclusions in steel, such as sulfur, tin, bismuth, and lead, can affect the sliding behavior of the mated materials. These inclusions act as solid lubricants. Thus, the sulfur-bearing stainless steels, such as types 303 and 416, have better galling resistance, but poorer adhesive wear resistance than their non-sulfur-bearing parent alloys, types 304 and 410. Another alloy example is Waukesha 88, which contains a tin- and bismuth-bearing second phase that results in excellent galling resistance, despite the high nickel content of the alloy (Ref 16).

Hardness and microstructure are critical to abrasive wear resistance. Unlike adhesive wear resistance, which favors austenitic alloys with a high work-hardening rate, abrasive wear resistance favors the hard martensitic matrix structure versus the softer austenitic or ferritic structure. This is especially significant in low-stress abrasion, where little subsurface deformation occurs. Also, high-carbon alloys with primary carbides have better wear resistance. Thus, martensitic alloy S44004 has good resistance to abrasion, whereas low-carbon martensitic PH alloy S17400 has poor abrasion resistance. For gouging abrasion, Hadfield Mn steel is used, because strain hardening and impact resistance are critical.

To resist fretting, thermal wear, and corrosive wear, corrosion and oxidation resistance become critical. The type of stainless steel to be used depends on the environment, with austenitic stainless steels favored in severe environments. A key factor for resisting these types of wear is the alloying addition, because silicon, aluminum, and chromium improve the corrosion and oxidation resistance of a particular alloy regardless of stainless family.

Wear and Galling Tests Commonly Used for Stainless Steels

Dry sand/rubber wheel test is an ASTM standard test (Ref 17) used to determine the resistance of metallic materials to low-stress (scratching) abrasion. The test involves the abrading of a test specimen with a grit of controlled size and composition (that is, rounded grain quartz sand). The abrasive is introduced between the test specimen and a rotating wheel with a chlorobutized rubber tire or rim of a specified hardness. The test specimen is pressed against the rotating wheel at a specified force by means of a lever arm while a controlled flow of grit abrades the test surface. The rotation of the wheel is such that its contact face moves in the direction of the sand flow. A schematic diagram of the test apparatus is shown in Fig. 6. The test duration and force applied by the lever arm vary depending on the relative wear resistance of the materials evaluated. For stainless steels, the test duration is one-third that of the more abrasion-resistant tool steel alloys such as D2. Specimens are weighed before and after the test, and the loss in mass is recorded. Abrasive wear is generally reported in terms of volume loss (in cubic millimeters) by dividing mass loss by the density of the alloy.

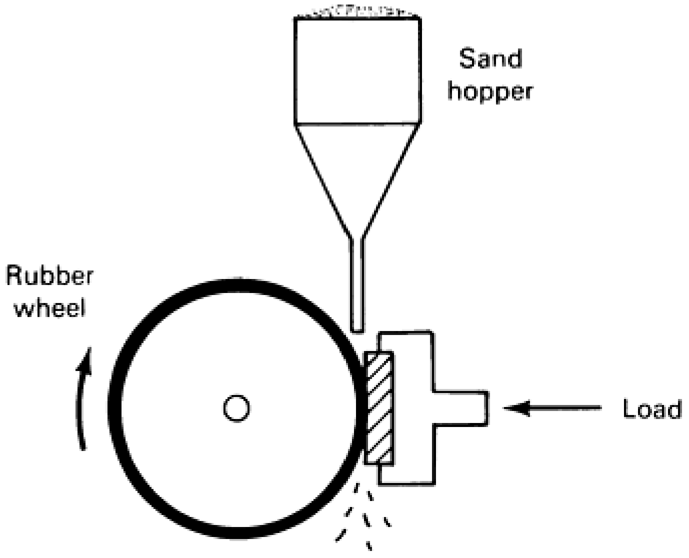


Fig. 6 Dry sand abrasion test apparatus. Source: Ref 18

Corrosive wear testing commonly involves hub and ball mill tests. The apparatus used for the hub test, which is low-stress abrasive test, is shown in Fig. 7. The slurry container has three hubs, each of which holds eight specimens, which are driven through slurry, as shown in Fig. 8.

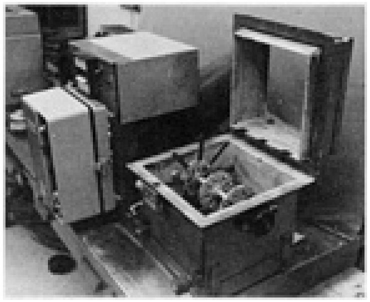


Fig. 7 Hub wear test machine. Source: Ref 12, p 3

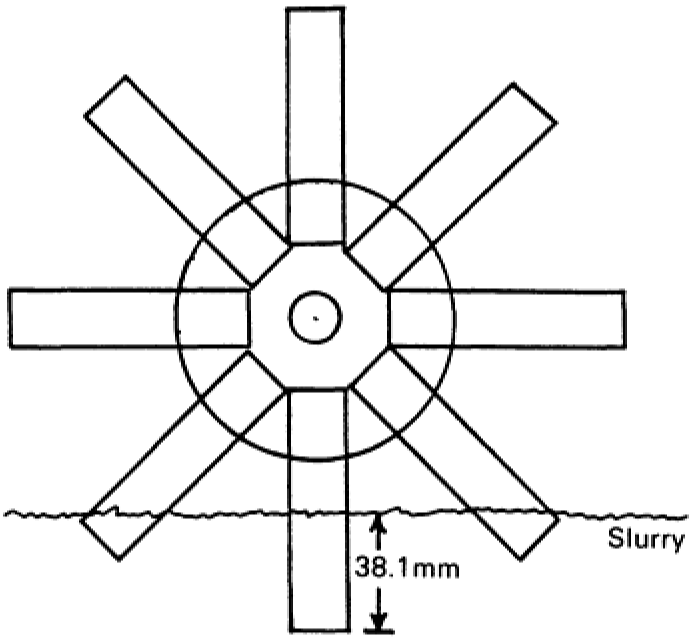


Fig. 8 Specimen arrangement on hub. Source: Ref 12, p 3

Slurries can differ, depending on the application that needs to be simulated. Often, alloys are evaluated both wet (corrosive slurry) and dry to demonstrate the effect of corrosion on corrosive wear. Specimens are weighed before and after the test, and volume loss is determined by dividing the weight loss by the alloy density.

The ball mill test is used extensively by the U.S. Bureau of Mines and in Canada to determine corrosive effects found in nickel mines. For this test, a 5.3 L (1.4 gal) porcelain jar serves as the ball mill, and the specimens are free to tumble in the corrosive slurry. Tests are performed for a particular time period, such as 8 h, after which the weight loss is determined and volume loss is calculated. The samples are then tested for additional time periods, using fresh slurry for each period. At the conclusion of the test, corrosive wear versus time is plotted for the alloys being evaluated.

Block-on-ring is an ASTM standard test (Ref 19) for determining the resistance of materials to sliding wear. The test utilizes a block-on-ring friction and wear testing machine to rank pairs of materials according to their sliding wear characteristics under various conditions. Rotational speed and load can be varied to correspond to service requirements. In addition, tests can be run with various lubricants, liquids, or gaseous atmospheres.

The test consists of a block specimen loaded against a ring specimen at a given speed for a given number of revolutions (Fig. 9). Block scar volume is calculated from the block scar width and depth, and ring scar volume is calculated from ring weight loss. The friction force required to keep the block in place is continuously measured during the test with a

load cell and is then recorded. The choice of test parameters is left to the user (with the exception of sliding distance, which is specified because wear does not usually vary linearly with distance in this test).

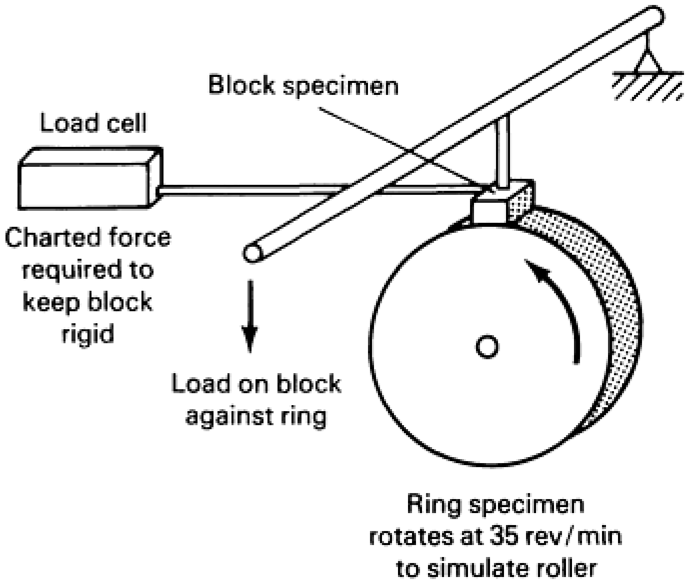


Fig. 9 Key components of a block-on-ring test apparatus. The coefficient of sliding friction, μ_{sf}, is calculated using the equation μ_{sf} = (friction force)/(applied load). Source: Ref 2

Crossed-cylinder is an ASTM standard test (Ref 20) for determining the resistance of metallic materials to metal-to-metal wear. This test ranks the adhesive wear resistance of materials and evaluates the compatibility of different metal couples. It is the most commonly used test for evaluating metal-to-metal wear resistance of stainless steels.

The test configuration consists of two cylindrical specimens that are positioned perpendicular to each other. One sample is rotated at a specified test speed while the other sample is kept stationary. The stationary specimen is pressed against the rotating specimen at a specified load by means of a lever arm and attached weights. The setup results in deadweight loading. A photograph of the test apparatus is shown in Fig. 10. Elevated-temperature tests have also been performed using the crossed-cylinder apparatus. The test duration, number of cycles, and rotational speed are varied depending on the relative wear resistance of the mated materials. The amount of wear is determined by weighing the specimen before and after the test. Weight loss is converted to volume loss by dividing the density of the material. Volume loss is determined for both the stationary and rotating specimens, and the total volume loss is recorded. When dissimilar materials are being tested, it is recommended that each alloy be tested in both the stationary and rotating positions.

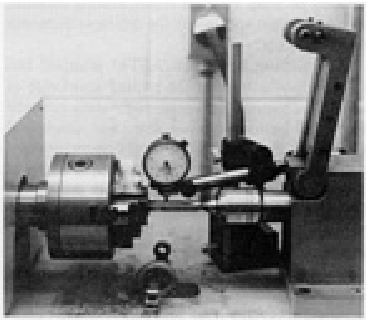


Fig. 10 Typical crossed-cylinder test apparatus. Source: Ref 21

Pin-on-disk is an ASTM standard test for determining the wear of material during sliding (Fig. 11). The coefficient of friction can also be determined using this test. For the pin-on-disk wear test, two specimens are required: one specimen is a pin with a radiused tip that is positioned perpendicular to the other specimen, which is usually a flat circular disk. A ball, rigidly held, often is used as the pin specimen. The test machine causes either the disk specimen or the pin specimen to revolve about the disk center. In either case, the sliding path is a circle on the disk surface. The plane of the disk may be oriented either horizontally or vertically. The pin specimen is pressed against the disk at a specified load, usually by means of an arm or lever and attached weights. Wear results are reported as volume loss (in cubic millimeters) for the pin and disk separately. When two different materials are tested, it is recommended that each material be tested in both the pin and disk positions.

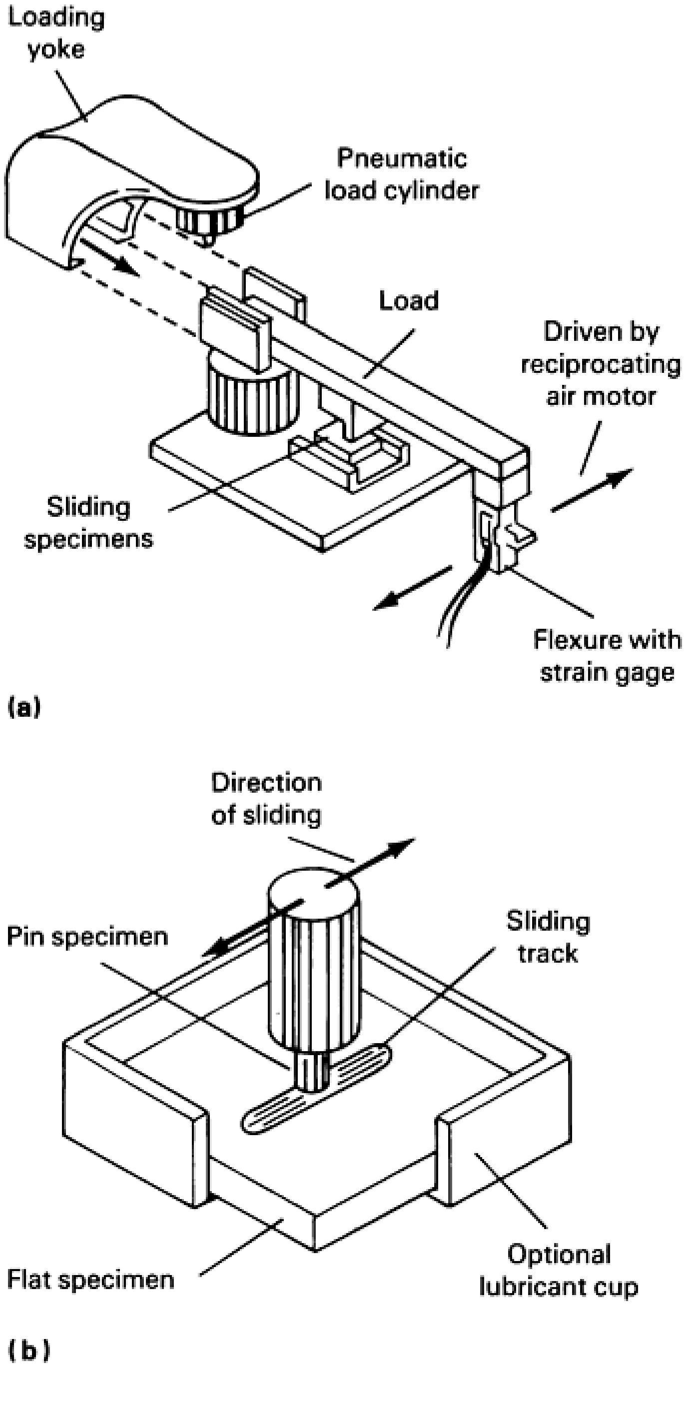


Fig. 11 Pin-on-disk wear test apparatus. (a) Key components of the device. (b) Close-up showing pin motion relative to flat when subjected to load. Source: Ref 2

The amount of wear is determined by measuring appropriate linear dimensions of both specimens before and after the test, or by weighing both specimens before and after the test. Linear measurements of wear are converted to wear volume by using appropriate geometric relationships, while mass loss is converted to volume loss by dividing mass loss by the appropriate density values for the specimens. Wear results are usually obtained by conducting a test for a selected sliding distance, load, and speed. Graphs of wear volume versus sliding distance using different specimens for different distances can be plotted.

The galling test is an ASTM standard test (Ref 22), which ranks the galling resistance of material couples. This test, commonly referred to as the button-on-block test, was developed during the 1950s and is the most commonly used procedure to test the galling resistance of stainless steels. Although it is generally performed on bare metals, nonmetallics, coatings, solid lubricants, and surface-modified alloys can be tested as well.

The test method uses available laboratory equipment that is capable of maintaining a constant compressive load between two flat specimens, such as hydraulic or screw-fed compression testing machines. A specimen with a 13 mm (0.5 in.) diameter section, referred to as the button, is slowly rotated one revolution (360°) relative to the other specimen under a specified compressive load. The test surfaces are ground so that the surface roughness range is from 0.4 to 1 μm (15 to 40 μin.) and the specimens are flat to ensure 100% contact between the mated surfaces.

The surfaces are examined for galling after sliding. The criterion for whether galling occurs is the appearance of the specimens, based on unassisted visual examination. Galling is characterized by at least one of the contacting surfaces exhibiting torn or raised metal. If the specimens have not galled, then a new button is tested on a new block location at an increased load. This procedure is continued until galling occurs. Similarly, if galling does occur on the first test, then lower loads are evaluated until no galling occurs. The galling test set-up is shown in Fig. 12.

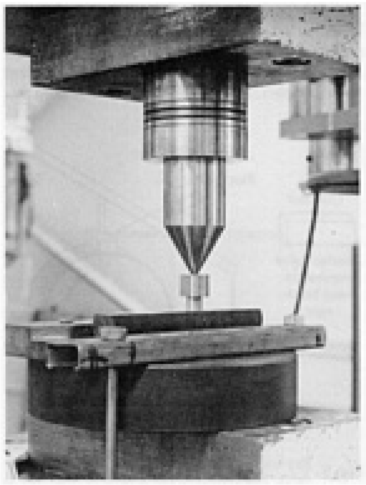


Fig. 12 Button-on-block galling test. Source: Ref 23

The loads applied correspond to a contact stress for the 13 mm (0.5 in.) diameter button. The stress midway between the highest nongalled test and the lowest galled test is referred to as the threshold galling stress. The higher the threshold galling stress, the more resistant the mated materials are to galling. Galling resistance can be determined for a particular selfmated material or dissimilar-mated materials. For the latter, it is recommended that each alloy be evaluated as button and block specimens.

To simulate repeated part performance under service conditions, such as those existing with sliding valve parts, the button specimen can be rotated multiple revolutions, in which the direction is reversed after each 360° revolution, instead of just rotating it once. Because of the increased severity of the multiple-rotation method, threshold galling stresses will be significantly lower than those of the single-rotation test.

To test highly resistant alloys, such as cobalt base Stellite 6B, a similar button-on-block test is used. The button is rotated through a 120° arc ten times at three different loads. Surface profilometry is then used to assess the degree of damage (Ref 24).

Another test that is similar to the button-on-block test is the ring-on-ring test, which uses ring specimens, rather than solid button/block samples. The test is used to assess galling resistance for tubular products related to oil production (Ref 25).

Galling of threaded parts, such as 9 m (30 ft) drill collars, is evaluated by a make/break test. The test procedure involves making a box/pin connection at a specified torque, breaking the connection, and recording the breakout torque. The procedure is repeated a given number of times, and the threads are periodically examined for galling. Unlike the button-on-block test, which simply tries to rank materials, this test simulates the actual service connections that are made for drill collars in the oil-drilling industry. Tests that simulate the specific application are very beneficial, but are usually quite expensive to perform. Thus, the button-on-block test can be used as a screening test to choose the best candidate alloy for a more specific test tailored to the service conditions of a particular application.

Wear Data for Stainless Steels

Abrasive/Corrosive Wear. The low-stress abrasion resistance of stainless steels can be determined by using either the dry sand abrasion or the hub test. Dry sand abrasion data (Table 4) clearly show that austenitic stainless steels have poorer wear resistance than the harder martensitic alloys, whose increased carbon content increases wear resistance. Type 410, which has 0.1% C, has wear resistance similar to the best austenitic alloys, whereas the wear resistance of type 440C is three to four times greater than the austenitic alloys. The abrasive wear resistance of type 440C is inferior to D2, an abrasive wear resistant tool steel. Test results shows that D2 has a volume loss of 32 mm^3, which is similar to type 440C, when the test time for D2 is tripled.

Table 4 Abrasive wear resistance of stainless steels

Alloy	Rockwell hardness	Volume loss, mm^3
Austenitic		
Type 304	85.5 HRB	113.1
S21061	95.0 HRB	139.6
S21061	26.5 HRC	138.2
S21800	93.0 HRB	127.2
S24100	95.5 HRB	94.9
S28200	97.0 HRB	88.3
Martensitic		
Type 410	44.0 HRC	94.0
S42010	51.0 HRC	67.5
Type 420	52.0 HRC	59.4
Type 440A	56.0 HRC	56.3
Type 440C	56.0 HRC	33.5
Type 440C	58.0 HRC	32.1

Note: Based on dry sand/rubber wheel abrasion test, ASTM G 65, procedure B, 133 N (30 lbf), 10 min duration.

Source: Ref 26

When a wet abrasive slurry is present, stainless steels experience greater wear than under dry conditions. However, their wear resistance is superior to martensitic carbon steels. Figure 13 compares five alloys under dry and wet slurry conditions in the hub test. Under dry conditions, the stainless alloys evaluated had similar or greater wear than Hadfield Mn steel and AISI 4340. However, when evaluated in a wet slurry of 5% NaCl plus 0.5% acetic acid, the stainless alloys were more resistant. Similar trends are shown in various wet environments, such as distilled water, synthetic nickel mine water, and synthetic seawater (Fig. 14, 15, 16) using the ball mill test.

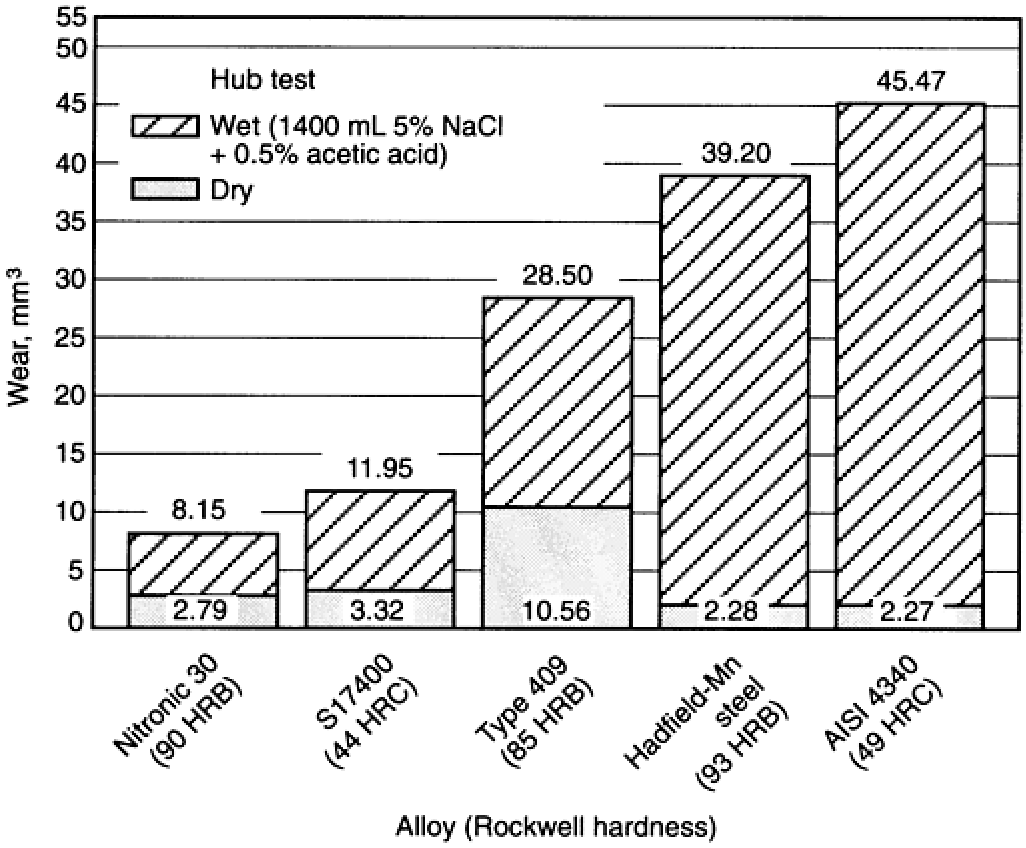


Fig. 13 Abrasive wear of alloy and stainless steels under dry and wet corrosive conditions. Source: Ref 12, p 4

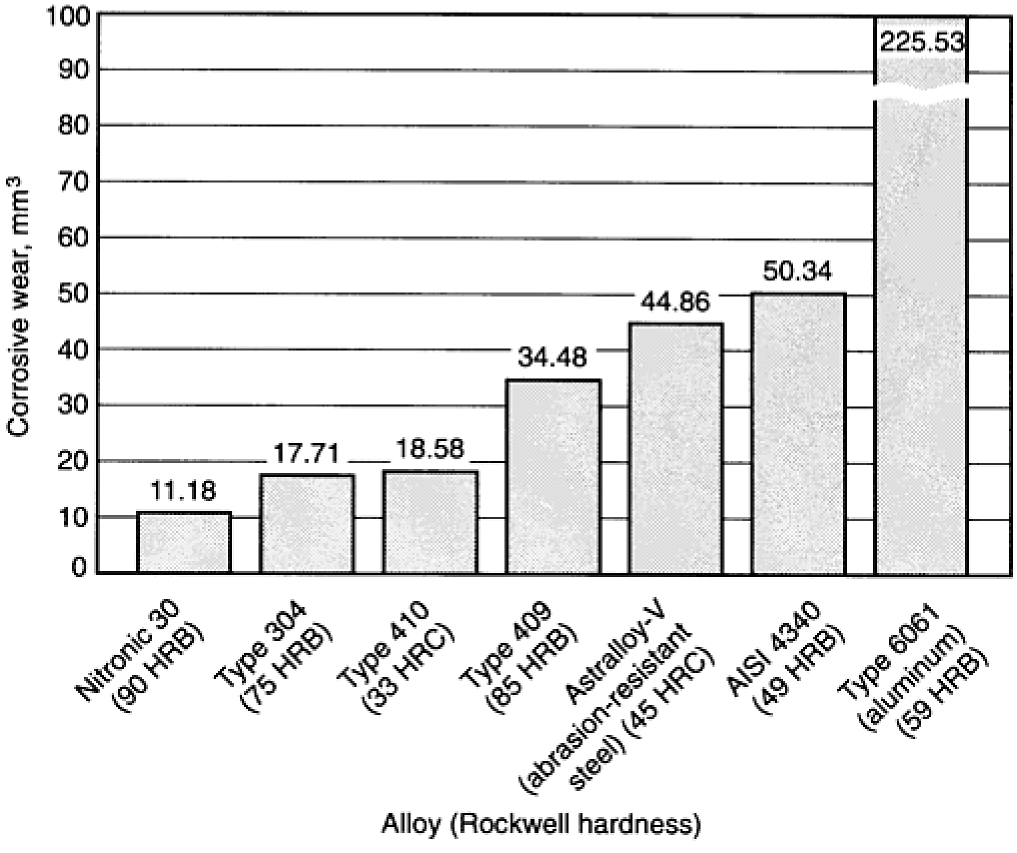


Fig. 14 Corrosive wear of alloy and stainless steels relative to an aluminum alloy, based on ball mill test using 1500 mL distilled water. Source: Ref 12, p 5

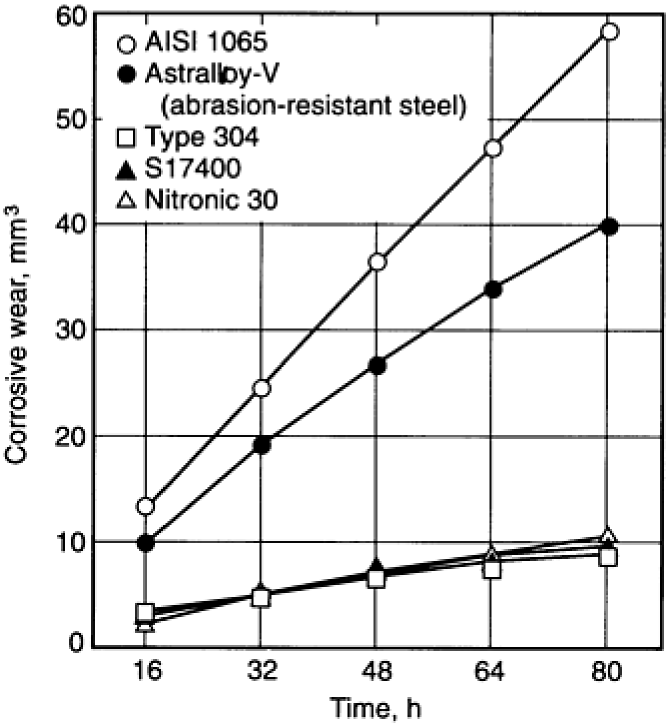


Fig. 15 Corrosive wear of alloy and stainless steels, based on ball mill test using synthetic nickel mine water. Source: Ref 12, p 8

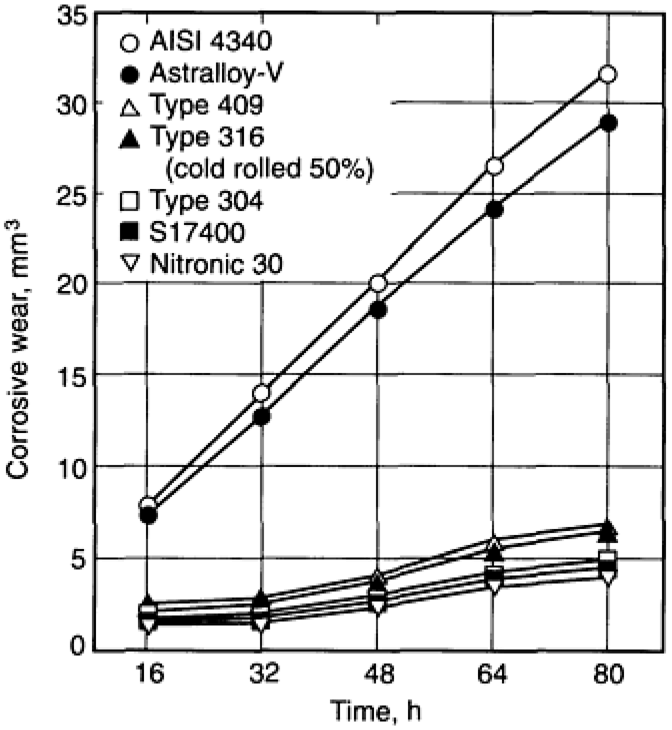


Fig. 16 Corrosive wear of alloy and stainless steels, based on ball mill test using synthetic seawater. Source:

Stainless steels have superior corrosive wear resistance in coal-handling equipment. Service life is greatly improved, compared with carbon steels (Fig. 17). In the mining industry, corrosion is the primary factor in the deterioration of equipment. Figure 18 shows the effect of chromium content on atmospheric corrosion resistance. Without chromium, corrosion rates are at least ten times greater than stainless steels.

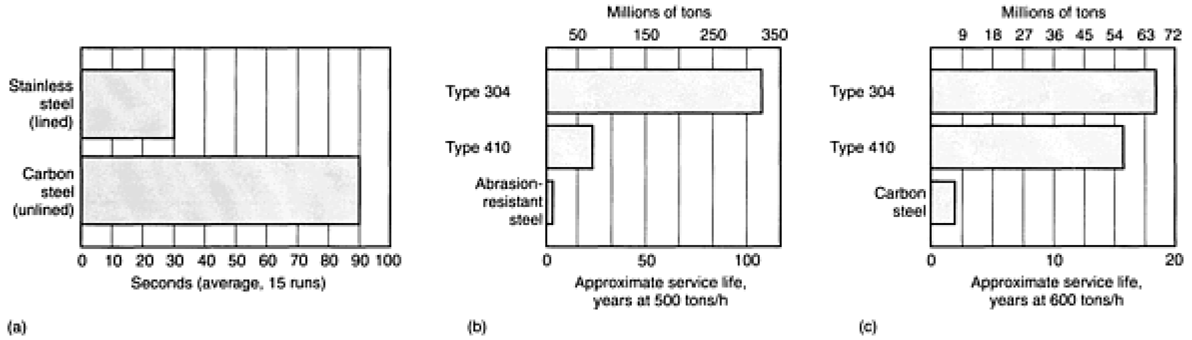


Fig. 17 Service life of coal-handling equipment. (a) Coal unloading time, 50 ton hopper car. (b) Coal conveyor bottom $\frac{3}{16}$ × 0 coal at 49 to 82 °C (120 to 180 °F). 6000 h/yr. (c) Coal conveyor bottom 5 × 0 clean coal. 6000 h/yr. Source: Ref 13

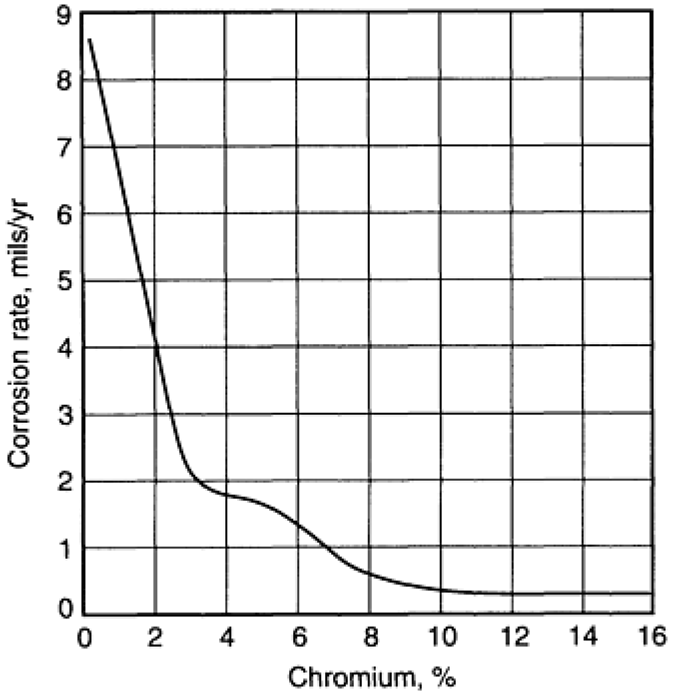


Fig. 18 Effect of chromium content on atmospheric corrosion resistance. Source:Ref 11

Tests under service conditions have been performed on stainless steels, abrasion-resistant steels, and carbon steels. Tests were performed for up to 5.3 years, and the reduction of thickness determined. Results of two studies show the superiority of stainless steels, compared with nonstainless alloys (Table 5). In general, type 304 is superior to type 410. However, alloy selection is dependent on the particular coal-handling equipment.

Table 5 Wear rates for test plates in drag conveyor bottoms

Material	Loss in μm/10^9 g (mils/million tons) of coal conveyed	Test time, months
125 × 9.5 mm (5 × $\frac{3}{8}$ in.)		
Type 304 stainless steel	0.67 (24)	27
Type 410 stainless steel	0.73 (26)	27
Type 316 stainless steel	0.90 (32)	27
Abrasion-resistant steel	1.01 (36)	15
Mild steel	1.65 (59)	15
9.5 mm ($\frac{3}{8}$ in.) by 0 thermal dryer product		
Type 304 stainless steel	0.045 (1.6)	25
Type 316 stainless steel	0.053 (1.9)	25
Type 410 stainless steel	0.100 (3.6)	25
Abrasion-resistant steel	2.3 (83)	6
63.5 × 6.35 mm (2.5 × 0.25 in.)		
Type 304 stainless steel	0.07 (2.6)	47
Type 410 stainless steel	0.11 (3.9)	47
Wrought iron	2.5 (89)	23
Mild steel	3.0 (107)	23
Crushed middlings conveyor		
Type 304 stainless steel	0.17 (6)	48
Abrasion-resistant steel	7 (250)	(failed in 18 months)

Source: Ref 11

Adhesive Wear. The metal-to-metal wear resistance of stainless steels can be determined by using the crossed-cylinder wear test. Unlike low-stress abrasion resistance, austenitic stainless steels generally have superior resistance, compared with martensitic stainless alloy (Table 6). The excellent wear resistance of type 201, type 301, S20161, and S21800 can be attributed to a high work-hardening rate. Additionally, the latter two alloys have a high silicon content, which not only increases the work-hardening rate, but results in a more adherent oxide film, thereby preventing the transition from mild oxidation wear to severe wear. Like abrasive wear resistance, a high hardness value is critical for martensitic stainless steels in order to achieve good adhesive wear resistance. For type 440C, high hardness (56 HRC) results in exceptional metal-to-metal wear resistance, whereas for type 440C (26 HRC), volume loss is high. Generally, high volume loss is characteristic of these alloys, even at 50 HRC.

Table 6 Adhesive wear resistance of stainless alloys

Alloy	Rockwell hardness	Volume loss, mm^3		
		10,000 cycles[(a)]**, at 100 rev/min**	**40,000 cycles**	
			at 100 rev/min	**at 400 rev/min**
Austenitic stainless steels				
Type 201	93.0 HRB	3.2	10.9	7.4
Type 304	78.0 HRB	10.4	23.5	22.0
Type 304	98.0 HRB	13.6	30.0	27.0
Type 304 HN	91.5 HRB	12.5	44.4	19.5
Type 304 HN	22.5 HRC	12.7	45.4	23.7
Type 301	90.0 HRB	6.9	. . .	. . .
Type 316	91.0 HRB	15.8	. . .	. . .
Type 303	98.0 HRB	488.7	. . .	. . .
S20161	96.0 HRB	1.3	5.3	4.4
S20161	30.0 HRC	1.5	5.5	4.5
S20910	96.0 HRB	11.7	46.0	16.7
S21000	88.0 HRB	12.0	43.8	9.9
S21800	93.0 HRB	4.2	12.1	5.2
S21900	92.0 HRB	11.9	45.0	34.1
S24100	95.5 HRB	7.2	21.9	5.2

S28200	99.0 HRB	8.1	27.9	2.8
S28200	35.0 HRC	7.5	22.5	2.7
S30430	72.0 HRB	18.2	42.8	29.1
N08020	87.5 HRB	22.6	84.3	28.2
Other stainless steels				
Type 410	24.5 HRC	114.0	544.0	251.1
Type 410	39.0 HRC	215.0	632.0	93.8
S42010	49.5 HRC	96.9	115.7	8.1
S42010	47.0 HRC	178.6	680.3	35.7
Type 420	50.5 HRC	83.8	84.0	8.5
Type 420	46.0 HRC	215.0	. . .	. . .
Type 440C	56.0 HRC	0.7	1.3	1.1
Type 440C	26.5 HRC	44.1	191.0	38.8
S17400	30.5 HRC	83.0	281.3	53.0
S17400	40.0 HRC	50.8	202.8	45.7
S13800	32.5 HRC	77.5	. . .	. . .
S13800	47.0 HRC	49.1	. . .	. . .
Type 431	42.0 HRC	235.6	. . .	. . .
Type 430	76.5 HRB	52.4	194.6	133.9
Type 430	80.5 HRB	58.5	229.7	171.6
S32950	23.0 HRC	15.8	50.5	22.5

Note: Based on crossed-cylinder wear test, ASTM G 83, 71 N (16 lbf).

Source: Ref 27

(a) Procedure A.

When different materials are mated together under load, wear testing is important for proper alloy selection. A general misconception for metal-to-metal wear and galling resistance is that a large difference in hardness between the two mated alloys is beneficial. Wear data suggest that alloys with good adhesive wear, such as type 440C (57 HRC) and S21800 (92 HRB), are excellent alloy selections, regardless of the hardness of the other alloy (Table 7).

Table 7 Wear compatibility of dissimilar-mated stainless steels

Alloy	Volume loss, mm^3						
	Type 304 (99 HRB)	**Type 316 (91 HRB)**	**S17400 (43 HRC)**	**S24100 (95 HRB)**	**S20910 (99 HRB)**	**S21800 (95 HRB)**	**Type 440C (57 HRC)**
Type 304	16.4	. . .	. . .	. . .	. . .	. . .	. . .
Type 316	13.5	16.4	. . .	. . .	. . .	. . .	. . .
S17400	31.7	23.7	67.7	. . .	. . .	. . .	. . .
S24100	10.8	12.1	22.1	9.5	. . .	. . .	. . .
S20910	11.5	12.2	20.1	10.6	12.8	. . .	. . .
S21800	7.7	5.5	6.9	4.1	4.5	3.6	. . .
Type 440C	5.3	5.0	14.6	4.0	5.5	3.1	4.9

Note: Based on crossed-cylinder wear test, ASTM G 83, 105 rev/min test speed, 10,000 cycles, 71 N (16 lbf); (weight loss/1000 cycles converted to total volume loss).

Source: Ref 28

The best stainless alloys have superior adhesive wear resistance, compared with aluminum (type 6061), abrasion-resistant steel (Astralloy-V), and AISI 4130 carbon steel. However, they are less resistant than Hadfield MN steel, D2 tool steel, and Stellite 6B, a hardfacing alloy (Table 8). These rankings can change when environmental factors such as corrosion, oxidation, and temperature are considered. For example, these three factors will have a much greater effect on alloys that contain little, if any, chromium. Even Stellite 6B, which does contain chromium, has reduced adhesive wear resistance at elevated temperatures (Table 9), unlike type 304 and S21800.

Table 8 Relative ranking of adhesive wear resistance of wrought materials

Wrought material	Rockwell hardness	Weight loss, mg/1000 cycles	
		at 105 rev/min	**at 415 rev/min**
D2 tool steel	61 HRC	0.46	0.34
AISI 4337	52 HRC	0.73	0.48
Stellite 6B	48 HRC	1.00	1.27
Hadfield Mn steel	95 HRB	1.25	0.41
Al (10.5%) bronze	87 HRB	2.21	1.52
Type 6061-T6 aluminum	59 HRB	17.06	21.15
AISI 4130	47 HRC	9.44	6.80
Astralloy V steel	46 HRC	213.58	8.22
Stainless alloy Type 440C	57 HRC	3.81	0.54
S21800	95 HRB	2.79	1.58
Type 201	90 HRB	4.95	4.68

Note: Based on crossed-cylinder wear test, ASTM G 83, 71 N (16 lbf), 10,000 cycles.

Source: Ref 29

Table 9 Effect of temperature on adhesive wear

Alloy	Room-temperature Rockwell hardness	Volume loss, mm^3				
		at 80 °C (175 °F)	**at 205 °C (400 °F)**	**at 315 °C (600 °F)**	**at 425 °C (800 °F)**	**at 540 °C (1000 °F)**
S21800	95 HRB	2.6	2.3	2.8	2.2	1.3
Type 304	79 HRB	14.7	13.5	14.5	13.2	. . .
Type 410	95 HRB	282.8	231.2	. . .	. . .	. . .
Type 410	40 HRC	42.1	. . .	. . .	. . .	. . .
S17400	35 HRC	31.3	23.8	12.5	8.2	. . .
Stellite 6B[(a)]	41 HRC	2.6	3.3	3.7	4.3	7.6

Note: Based on crossed-cylinder test, ASTM G 83, using test speed of 415 rev/min for 20,000 cycles, 71 N (16 lbf).

Source: Ref 30

(a) Cobalt-base hardfacing alloy.

The galling resistance of stainless steels can be determined by using the button-on-block test. Higher threshold galling stress indicates better resistance to galling. All five families of stainless steels are susceptible to galling.

For austenitic stainless steels, Cr-Mn-N alloys generally have superior galling resistance, compared with the chromium-nickel alloys (Table 10, 11). Nickel is considered harmful to galling resistance, whereas silicon is considered beneficial. Thus, S20161 and S21800 have excellent resistance. Increasing the hardness of austenitic stainless alloys by cold drawing decreases galling resistance.

Table 10 Threshold galling stress results for selected self-mated stainless steels

Stainless steel	Condition	Rockwell hardness	Threshold galling stress	
			MPa	**ksi**
Austenitic				
S20161	Annealed	95 HRB	104[(a)]	15[(a)]
S21800	Annealed	92 HRB	104[(a)]	15[(a)]
S28200	Annealed	96 HRB	166	24
S28200	Cold drawn	35 HRC	62	9
S20910	Annealed	97 HRB	35	5
S20900	Annealed	96 HRB	48	7
S24100	Annealed	23 HRC	97	14
S30430	Annealed	74 HRB	35	5
Type 304	Annealed	86 HRB	55	8
Type 304	Cold drawn	27 HRC	17	2.5
Type 316	Annealed	82 HRB	48	7
Type 316	Cold drawn	27 HRC	35	5
Type 303	Annealed	85 HRB	138	20
Type 201	Annealed	95 HRB	104	15
N08020	Annealed	87 HRB	14	2
Martensitic				
S42010	Tempered at 204 °C (400 °F)	50 HRC	104[(a)]	15[(a)]
S42010	Tempered at 260 °C (500 °F)	47 HRC	62	9
Type 410	Annealed	87 HRB	7	1
Type 410	Tempered at 260 °C (500 °F)	43 HRC	21	3
Type 416	Annealed	95 HRB	21	3
Type 416	Tempered at 316 °C (600 °F)	37 HRC	62	9
Type 416	Tempered at 538 °C (1000 °F)	32 HRC	42	6
Type 420	Tempered at 204 °C (400 °F)	51 HRC	25	18
Type 420	Tempered at 260 °C (500 °F)	49 HRC	55	8
Type 440C	Tempered at 260 °C (500 °F)	55 HRC	125	18
Precipitation hardenable				
S45500	Aged at 510 °C (950 °F)	48 HRC	90	13
S45500	Aged at 566 °C (1050 °F)	43 HRC	59	8.5
S45500	Aged at 621 °C (1150 °F)	36 HRC	28	4
S45000	Annealed	29 HRC	69	10
S45000	Aged at 480 °C (895 °F)	43 HRC	55	8
S45000	Aged at 566 °C (1050 °F)	38 HRC	17	2.5
S45000	Aged at 621 °C (1150 °F)	33 HRC	14	2
S17400	Aged at 480 °C (895 °F)	45 HRC	69	10
S17400	Aged at 621 °C (1150 °F)	34 HRC	35	5
S13800	Aged at 538 °C (1000 °F)	46 HRC	21	3
S66286	Aged at 718 °C (1325 °F)	30 HRC	14	2
Ferritic and duplex				
S18200	Cold drawn	98 HRB	35	5
Type 430F	Annealed	92 HRB	14	2
Type 430	Cold drawn	98 HRB	10	1.5
Type 329	Annealed	25 HRC	7	1

Note: Based on button-on-block test, ASTM G 98, unlubricated ground finish.

Source: Ref 31

(a) Did not gall.

Table 11 Threshold galling stress for selected stainless steels

Alloy	Condition and nominal Rockwell hardness	Type 410		Type 416		Type 430		Type 440C		Type 303		Type 304		Type 316		S17400		S24100		S21800	
		MPa	**ksi**	**MPa**	**ksi**	**MPa**	**ksi**	**MPa**	**ksi**	**MPa**	**ksi**	**MPa**	**ksi**	**MPa**	**ksi**	**MPa**	**ksi**	**MPa**	**ksi**	**MPa**	**ksi**
Type 410	Tempered (38 HRC)	21	3	28	4	21	3	21	3	28	4	14	2	14	2	21	3	317	46	345 [a]	50
Type 416	Tempered (36 HRC)	28	4	90	13	21	3	145	21	62	9	165	24	290	42	14	2	310	45	345 [a]	50
Type 430	Annealed (84 HRB)	21	3	21	3	14	2	14	2	14	2	14	2	14	2	21	3	55	8	248	36
Type 440C	Tempered (56 HRC)	21	3	145	21	14	2	76	11	34	5	21	3	255	37	21	3	345 [a]	50	345 [a]	50
Type 303	Annealed (82 HRB)	28	4	62	9	14	2	34	5	14	2	14	2	21	3	21	3	345 [a]	50	345 [a]	50
Type 304	Annealed (77 HRB)	14	2	165	24	14	2	21	3	14	2	14	2	14	2	14	2	207	30	345 [a]	50
Type 316	Annealed (81 HRB)	14	2	290	42	14	2	255	37	21	3	14	2	14	2	14	2	21	3	262	38
S17400	Aged (84 HRB)	21	3	14	2	21	3	21	3	14	2	14	2	14	2	14	2	345 [a]	50	345 [a]	50
S21800	Annealed (94 HRB)	345 [a]	50	345 [a]	50	248	36	345 [a]	50	345 [a]	50	345 [a]	50	262	38	345 [a]	50	345 [a]	50	345 [a]	50

Note: Based on button-on-block test, ASTM G 98, unlubricated ground finish; condition and hardness apply to both horizontal and vertical axes.

Source: Ref 28

(a) Did not gall.

For the heat-treat-hardenable stainless steels, increasing hardness increases galling resistance. For martensitic PH stainless steels, increasing aging temperature may decrease galling resistance. For example, tests on S13880 and S17440 with similar hardness values (45 to 46 HRC) resulted in a much lower threshold galling stress (21 versus 69 MPa, or 3 versus 10 ksi) for S13800, which was aged at a higher temperature. Martensitic alloys, such as type 440C, tend to have heavy scoring damage, rather than galling.

Button-on-block test severity can be increased by increasing the number of button rotations against the block (Table 12). A multiple-rotation test more closely simulates the action of valve components. The threshold galling stress values for alloys such as type 303, S28200, S42010, S24100, S45500, and type 440C were low (<14 MPa, or 2 ksi), and were greatly reduced, compared with the single-rotation values. For high-silicon alloys, such as S20161 and S21800, improved galling resistance is clearly demonstrated using the triple-rotation test.

Table 12 Threshold galling stress results for selected self-mated stainless steels

Alloy	Rockwell hardness	Single rotation[a]		Triple rotation	
		MPa	**ksi**	**MPa**	**ksi**
S20161	95 HRB	104[b]	15[b]	104[b]	15[b]
S20161	28 HRC	104[b]	15[b]	104[b]	15[b]
S28200	96 HRB	166	24	7	1
S21800	92 HRB	104[b]	15[b]	48	7
T440C	55 HRC	124	18	14	2
T304	86 HRB	55	8	7[c]	1[c]
T430	98 HRB	10	1.5	7[c]	1[c]
S42010	50 HRC	104[b]	15[b]	21	3
T420	49 HRC	55	8	14	2
S24100	23 HRC	97	14	14	2
S45500	48 HRC	97	13	7[c]	1[c]
S66286	30 HRC	14	2	7[c]	1[c]
Type 303	85 HRB	138	20	7[c]	1[c]

Note: Based on button-on-block test, unlubricated ground finish.

Source: Ref 14, p 65, Ref 32

(a) Per ASTM G 98.

(b) Did not gall.

(c) Galled at this stress level.

Dissimilar-mated test results (Tables 11 and 13) show the need to screen dissimilar-mated couples prior to service. Some dissimilar couples have threshold galling stress (TGS) values that are higher than either alloy when self-mated and vice versa. For example, type 304 mated with S45500 has a TGS of 124 MPa (18 ksi), whereas their self-mated values are 55 and 90 MPa (8 and 13 ksi), respectively. Type 304 mated with type 440C has a TGS of 28 MPa (4 ksi), whereas self-mated values are 55 and 124 MPa (8 and 18 ksi), respectively.

Table 13 Threshold galling stress results for selected dissimilar-mated stainless steels

Alloy and hardness	Stress	
	MPa	**ksi**
Type 201 (94 HRB)		
versus type 304 (77 HRB)	14	2
versus S17400 (41 HRC)	14	2
versus S24100 (21 HRC)	284	36
Type 301 (86 HRB)		
versus type 416 (35 HRC)	21	3
versus type 440C (56 HRC)	21	3
versus S21800 (95 HRB)	345[a]	50[a]
Type 304 (86 HRB)		
versus type 304 (27 HRC)	28	4
versus type 440C (55 HRC)	28	4
versus S28200 (96 HRB)	41	6
versus S20910 (97 HRB)	69	10
versus S24100 (23 HRC)	104[a]	15[a]
versus S45000 (43 HRC)	21	3
versus S45500 (48 HRC)	124	18
Type 304 (77 HRB)		
versus S17400 (33 HRC)	14	2
versus S17400 (46 HRC)	14	2
Type 316 (82 HRB)		
versus type 316 (27 HRC)	55	8
versus type 440C (55 HRC)	14	2

Type 316 (27 HRC)		
versus type 329 (25 HRC)	14	2
S30430 (74 HRB)		
versus type 440C (55 HRC)	83	12
S28200 (35 HRC)		
versus S28200 (96 HRB)	104	15
S28200 (96 HRB)		
versus S45000 (29 HRC)	55	8
S28200 (96 HRB)		
versus S45000 (43 HRC)	62	9
S24100 (22 HRC)		
versus S17400 (39 HRC)	76	11
S24100 (22 HRC)		
versus S20910 (34 HRC)	55	8
S24100 (43 HRC)		
versus S20910 (95 HRB)	90	13
S21800 (95 HRB)		
versus S20910 (95 HRB)	345[a]	50[a]
versus S66286 (28 HRC)	345[a]	50[a]
versus type 420 (50 HRC)	345[a]	50[a]
versus S17400 (33 HRC)	345[a]	50[a]
versus S13800 (44 HRC)	345[a]	50[a]
S21800 (92 HRB)		
versus N08020 (87 HRB)	48	7
Type 410 (42 HRC)		
versus type 440C (55 HRC)	35	5
versus S45000 (43 HRC)	7	1
Type 410 (32 HRC)		
versus type 420 (50 HRC)	21	3
versus type 416 (38 HRC)	28	4
Type 416 (32 HRC)		
versus type 416 (83 HRB)	76	11
Type 416 (37 HRC)		
versus type 440 (55 HRC)	159	23
Type 416 (34 HRC)		
versus type 430 (90 HRB)	21	3
S45500 (48 HRC)		
versus S17400 (45 HRC)	76	11
S45500 (48 HRC)		
versus S17400 (38 HRC)	55	8
S13800 (46 HRC)		
versus S17400 (45 HRC)	62	9
S13800 (46 HRC)		
versus S17400 (38 HRC)	14	2
S17700 (44 HRC)		
versus S17400 (44 HRC)	14	2
S17700 (41 HRC)		
versus S17400 (41 HRC)	21	3

Note: Based on button-on-block test, ASTM G 98, unlubricated ground finish.

Source: Ref 28

(a) Did not gall.

In both examples, there is a large difference in hardness between the two members of these couples, but the resultant threshold galling stress values are quite different. For one case, the value is greater than the self-mated values, and for the other, it is lower.

Applications

Various applications for stainless steels can be found where parts are subjected to wear conditions and require the corrosion resistance that stainless steels offer. Several applications have already been discussed, such as coal-handling equipment, where the abrasivity of coal combines with the corrosivity of a wet slurry.

In food-processing equipment, lubricants are prohibited because of sanitation requirements, which means that metal-to-metal sliding can result in wear. For example, most food or drug products must be conveyed down a process line. The roller and side plate interface of a continuous chain-link conveyor belt results in sliding contact. Fine sugar, for example, can be entrapped in sliding parts and further enhances wear. In the food/drug processing industry, fasteners also must be used without the benefit of lubricants. Pump and valve applications are also affected by the sanitation required.

In the exploration for oil and gas, the corrosive environments encountered necessitate the use of stainless steel. For example, drill collars used in oil drilling have threaded connections that are often not effectively lubricated and are prone to galling. Valve parts, often under high pressure and loads, can be prone to galling and thus become ineffective. Leaky valves require repair and costly downtime.

The automotive industry uses programmed valves that redirect the hot exhaust gases away from the carburetor/manifold. A stainless steel was chosen as a stem material because of the elevated temperature (815 °C, or 1500 °F) encountered. In this environment, galling resistance is critical for the valve to last a minimum of 80,500 km (50,000 mi).

In structural fastener applications, corrosion failures of pin-and-hanger bridge assemblies have occurred while using carbon steel. Because of the load of the girders on the pin and hanger assembly, galling and metal-to-metal wear are important concerns; to avoid corrosion failures, however, stainless alloys are being evaluated (Ref 33).

Galling occurred on yacht turnbuckles when silicon bronze was used. Currently in use are stainless steels that have both seawater corrosion and galling resistance.

Because of their longer service life, stainless steels, such as type 304, 3Cr12, and type 410, are replacing carbon steel in corrosive wear applications. The stainless alloys that are more galling and wear resistant, such as S20161 and S21800, are being considered as replacements for Stellite 6B, a cobalt-base alloy, in valve and pump applications, because of their lower cost.

In the nuclear industry, alloys containing high levels of cobalt are being scrutinized, because wear and corrosion products can be released into the primary cooling water and be transported to the reactor core, where Co-59 is transmuted to the radioactive isotope Co-60. Noncobalt-containing stainless steels that are galling and wear resistant are being considered for this application. A family of alloys called NOREM has been developed by the Electric Power Research Institute, and are designed to replace cobalt-containing hardfacing alloys (Ref 34). Several parts that can experience wear and galling problems are shown in Fig. 19.

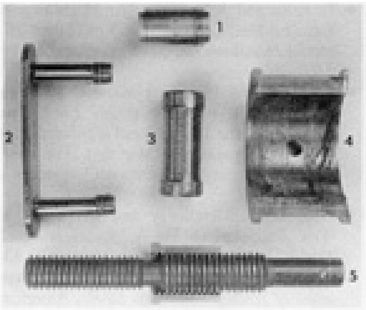


Fig. 19 Parts subject to wear and galling problems. (1) Bushing for chain belt. (2) Link for chain belt. (3) Bushing for chain. (4) Rod end bearing. (5) Valve stem. Source: Ref 28

Design Considerations

For applications in which metal-to-metal wear and galling are of concern, these guidelines should be followed:

- Lubricate where possible
- Keep load, temperature, and speed as low as possible
- Parts should be dimensionally tight with sufficient clearance
- Use a surface finish between 0.25 and 1.75 μm (10 and 70 μin.) whenever possible (many stainless parts are electropolished, which can lead to galling and wear)
- Increase contact area, so that there is less stress on parts and less depth of wear
- Carefully select alloys in unlubricated systems, or where insufficient lubrication may be present. Dissimilar-mated couples with high threshold galling stress values can be chosen or high work-hardening rate austenitic stainless alloys can be selected for improved adhesive and cavitation wear resistance and galling resistance
- Use surface treatments, such as nitriding, hard-face coating, shot peening, and carburizing, which are especially effective in combating fatigue wear

For applications in which either abrasive or corrosive wear is of concern:

- Use a high-hardness martensitic stainless steel when abrasive wear is involved
- Design equipment so that abrasive material flows uniformly
- Minimize the damage that is due to corrosion by designing equipment to minimize standing water or slurry
- Consider the use of stainless alloys when abrasive slurries are encountered

References

1. W.J. Schumacher, Developing New Answers to Galling and Wear, *SAMPE J.,* May/June 1977, p 16-19
2. J.H. Magee, Silicon Beefs Up Stainless Steel, *Mach. Des.,* Oct 1990, p 60-64
3. *Metals and Alloys in the Unified Numbering System,* 4th ed., Society of Automotive Engineers and ASTM, 1986
4. C.J. Novak, Wrought Austenitic Stainless Steels, *Handbook of Stainless Steels,* D. Peckner and I.M. Bernstein, Ed., McGraw-Hill, 1977, p 4-34
5. *Selecting Carpenter Stainless Steels,* Carpenter Technology Corporation, 1987, p 149, 182, 222, 223
6. Data sheets on aluminum 6061, filing code Al-205; data sheets on AISI 1080, filing code Cs-95; data sheets on Ambraloy-901 (aluminum bronze, 5% Al), filing code Cu-73, *Alloy Digest,* Jan 1973, July 1983, Apr 1959, respectively
7. *Review of the Wear and Galling Characteristics of Stainless Steels,* Committee of Stainless Steel Producers, American Iron and Steel Institute, Apr 1978
8. F. Borik, Metallurgy of Ferrous Materials for Wear Applications, reprinted from *Wear Control Handbook,* American Society of Mechanical Engineers, p 327 -342
9. Hard-Facing Materials, *Metals Handbook,* Vol 3, 9th ed., American Society for Metals, 1980, p 564 -565
10. E.I. Meletis, Wear Corrosion Processes in Ion-Plated Thin Films, *Tribological Mechanisms and Wear Problems in Materials,* Proceedings from ASM Materials Week '87, Oct 1987, ASM International, p 61-70
11. Committee of Stainless Steel Producers, *Stainless Steel: Effective Abrasion and Corrosion Control in Coal Handling and Preparation Equipment,* American Iron and Steel Institute, Apr 1976
12. W.J. Schumacher, Nitrogen Strengthened Austenitic Stainless Steel for Improved Corrosive Wear Resistance, *New Developments in Stainless Steel Technology,* American Society for Metals Conference

Proceedings, 1985, p 107-116
13. W.J. Schumacher, "Metals for Nonlubricated Wear," Armco Stainless Steel Products, reprint from *Mach. Des.,* 1976
14. P. Crook and A. Asphahani, Alloys to Protect against Corrosion and Wear, *Chem. Eng.,* 10 Jan 1983
15. L.K. Ives, M.B. Peterson, and E.P. Whitenton, "Galling: Mechanism and Measurement," National Bureau of Standards Report, p 33-40
16. D.K. Subramanyam and J. Hoag, Anti-Galling Alloy with Bismuth, *The Bulletin of the Bismuth Institute,* Bismuth Institute, 1985, p 1-12
17. "Standard Practice for Conducting Dry Sand/Rubber Wheel Abrasion Test," G 65, *Annual Book of ASTM Standards,* Vol 03.02, ASTM
18. S.J. Matthew, R.D. Zordan, and P. Crook, Laboratory Engineers Solutions to Unlubricated Wear Problems, *Met. Prog.,* July 1984, p 63
19. "Standard Practice for Ranking Resistance of Materials to Sliding Wear Using Block-on-Ring Wear Test," G 77 *Annual Book of ASTM Standards,* Vol 03.02, ASTM
20. "Test Method for Wear Testing with a Crossed-Cylinder Apparatus," G 83,*Annual Book of ASTM Standards,* Vol 03.02, ASTM
21. J.H. Magee, "Evaluation of Metal-to-Metal Crossed Cylinder Wear Test for Stainless Steels," Carpenter Technology Corporation, 1980
22. Standard Test Method for Galling Resistance of Materials, G 98, *Annual Book of ASTM Standards,* Vol 03.02, ASTM
23. J.H. Magee, Austenitic Stainless Steel with Improved Galling Resistance, *High Manganese Austenitic Steel,* Proceedings from ASM Materials Week '87, Oct 1987, ASM International, p 62
24. R.W. Kirchner, P. Crook, and A.I. Asphahani, "Wear/Corrosion-Resistant, High Performance Alloys for the Food Industries," Paper 102, presented at Corrosion 1984, National Association of Corrosion Engineers
25. H.A. Domian *et al.,* "Austenitic Fe-Cr-Ni Alloy Designed for Oil Countries Tubular Products," U.S. patent 4,840,768, 20 June 1987, p 5-6
26. "Abrasive Wear Data on Stainless Steels," Carpenter Technology Corporation
27. "Adhesive Wear Data on Stainless Steels," Carpenter Technology Corporation
28. Committee of Stainless Steel Producers, *Review of the Wear and Galling Characteristics of Stainless Steels,* American Iron and Steel Institute, Apr 1978
29. "Armco Nitronic 60 Stainless Steel (UNS 21800)," Armco product data bulletin NOS-45, 1984
30. W.J. Schumacher, High Temperature Wear of Stainless Steels, *Mech. Eng.,* Mar 1985, p 23-25
31. Galling and Stainless Steels--How to Overcome Galling, *Technical Department Bulletin,* Vol 1 (No. 11), Carpenter Technology Corporation
32. "Galling Wear Data on Stainless Steels," Carpenter Technology Corporation
33. W.J. Schumacher, Nitronic 60--The First Decade, *Proceedings of ASM Materials Week '87,* Oct 1987
34. H. Ocken, Implementing Cobalt-Free Alloys in Nuclear Plant Valves, *Radiation Control News,* EPRI-No. 6, Electric Power Research Institute, June 1990

Selected References

• C. Allen, B.E. Protheroe, and A. Ball, The Abrasive-Corrosive Wear of Stainless Steels, *Wear of Materials 1981,* Proceedings of the International Conference on Wear of Materials (San Francisco), 30 Mar to 1 Apr 1981, American Society of Mechanical Engineers, p 271-279
• K.C. Barker and A. Ball, Synergistic Abrasive-Corrosive Wear of Chromium Containing Steels, *Br. Corros. J.,* Vol 24 (No. 3), p 222-228
• R.C. Bill, *Review of Factors That Influence Fretting Wear, Materials Evaluation Under Fretting Conditions,* STP 780, ASTM, June 1981, p 165-182
• R.M. Dawson, T.A. DeBold, and M.J. Johnson, Corrosion of Stainless Steels, *Corrosion,* Vol 13, 9th

ed., *Metals Handbook,* ASM International, 1987, p 547-565

- E.D. Doyle and S.K. Dean, Abrasive Wear of Austenitic Stainless Steel, *Metal Forum,* Vol 4 (No. 4), 1981, p 235-245
- C.J. Heathcock and B.E. Protheroe, Cavitation Erosion of Stainless Steels, *Wear,* Vol 81, 1982, p 311-327
- K.L. Hsu, T.M. Ahn, and D.A. Rigney, Friction, Wear, and Microstructure of Unlubricated Austenitic Stainless Steels, *Wear of Materials 1979,* Proceedings of the International Conference on Wear of Materials, Apr 1979, American Society of Mechanical Engineers, p 12-26
- L.K. Ives, M.B. Peterson, and E.P. Whitenton, Mechanism of Galling and Abrasive Wear, *Fossil Energy Materials Program Conference Proceedings,* Aug 1987, p 397-421
- T. Kosa and R.P. Ney, Sr., Machining of Stainless Steels, *Machining,* Vol 16, 9th ed., *Metals Handbook,* ASM International, 1989, p 681-707
- K.C. Ludema, Selecting Material for Wear Resistance, *Wear of Materials 1981,* Proceedings of the International Conference on Wear of Materials, 30 Mar to 1 Apr 1981, American Society of Mechanical Engineers, p 1-4
- R.A. Lula, Metallurgy and Properties of Wrought Stainless Steels, *Stainless Steel,* American Society for Metals, 1986, p 29-88
- J.G. Parr and A. Hanson, *Mechanical Properties of Wrought Stainless,* American Society for Metals, 1965, p 25-68
- D. Peckner and I.M. Bernstein, Ed., *Handbook of Stainless Steels,* McGraw-Hill, 1977
- E. Rabinowicz, Adhesive Wear Values as Affected by Strength Fluctuations, *Wear of Materials,* 1981, p 197-201
- G. Schiefelbein, Performance of Alloys Against Erosion-Corrosion Attack, *Mater. Protect.,* Vol 9 (No. 6), June 1970, p 11-13
- W.J. Schumacher, Adhesive Wear Resistance of Engineering Alloys, *Met. Prog.,* Nov 1978, p 32-36
- W.J. Schumacher, Wear and Galling of Nitrogen-Strengthened Stainless Steels, *Mach. Des.,* Aug 1983, p 87-88

Friction and Wear of Bearing Steels

Charles A. Moyer, The Timken Company

Introduction

REGARDING WEAR for rolling bearing steels, it is important to identify the wear mechanisms that are primarily of concern. Using the mechanisms proposed by Vingsbo (Ref 1), metal loss in bearings is considered to come primarily from abrasive wear, adhesive wear, and contact fatigue wear. Within a properly operating rolling-element bearing, there is generally a minimal amount of sliding that takes place on the primary contact surfaces. The main wear mechanism is contact fatigue. Normally, wear is considered to be accompanied by substantial plastic deformation on or just below the surface, and this is typically the case with abrasive or adhesive wear.

For contact fatigue without gross sliding, the plastic flow that occurs is more localized and can be below what is normally considered surface-related wear and fracture. The depths of maximum shear (orthogonal shear, octahedral shear, or von Mises stress parameter) occur below a rolling contact surface with little surface sliding, tangential force, or surface asperity interaction. This maximum shear depth is on order of 1/50 of the bulk cross section of various bearings. Thus, the whole stress field from a Hertzian contact surface to below any localized maximum subsurface stresses can be defined as the active wear layer for rolling contacts. In this article, bearing steel wear will be considered to include contact fatigue, and the total material involved will range from the surface topography to ~0.1 to 1 mm (~0.004 to 0.04 in.) below the surface. Wear distance for bearing steel under high-cycle fatigue is usually measured within a bearing by how far a point

on one of the bearing raceways has traveled. This distance is usually expressed in total revolutions of one of the raceways or as a percentage of the bearing's expected life.

Composition of Bearing Steels

The steel used in rolling bearings for normal requirements such as industrial or automotive applications is an alloy steel with alloy content, in addition to carbon, that runs from ~1.5 to 6%, depending on the bearing ring cross section and hardenability requirements. However, bearing steels are also known to have reduced non-metallic inclusions so that rolling contact fatigue life requirements can be met. Typical standard bearing steel compositions for high-carbon or through-hardened steels are given in Table 1, and standard bearings steel compositions for low-carbon or carburizing bearing steels are given in Table 2 (Ref 2). Both steels are use depending on specific service condition needs. The advantages of each are pointed out by Burrier (Ref 2). High-carbon steels provide the following advantages:

- Ability to carry somewhat higher contact stresses, such as those encountered in point contact loading in ball bearings
- A simpler quench and temper heat treatment, when compared to carburizing
- Potentially greater dimensional stability under temperature extremes because of their characteristically lower content of retained austenite

Carburizing steels, on the other hand, offer the following advantages:

- Greater surface ductility (due to their retained austenite content) to better resist the stress-raising effects of asperities, misalignment, and debris particles
- A higher level of core toughness to resist through-section fracture under severe service conditions
- A compressive residual surface stress condition to resist bending loads imposed on the ribs of roller bearings, and to reduce the rate of fatigue crack propagation through the cross section
- Easier machining of the base material in manufacturing

Table 1 Nominal compositions of high-carbon bearings steel

Grade	Composition, %					
	C	**Mn**	**Si**	**Cr**	**Ni**	**Mo**
AISI 52100	1.04	0.35	0.25	1.45	. . .	. . .
ASTM A 485-1	0.97	1.10	0.60	1.05	. . .	. . .
ASTM A 485-3	1.02	0.78	0.22	1.30	. . .	0.25
TBS-9	0.95	0.65	0.22	0.50	0.25 max	0.12
SUJ 1[(a)]	1.02	<0.50	0.25	1.05	<0.25	<0.08
105 Cr6[(b)]	0.97	0.32	0.25	1.52	. . .	. . .
SHKH15-SHD[(c)]	1.00	0.40	0.28	1.48	<0.30	. . .

Source: Ref 2

(a) Japanese grade.

(b) German grade.

(c) Russian grade.

Table 2 Nominal compositions of carburizing bearings steels

Grade	Composition, %					
	C	**Mn**	**Si**	**Cr**	**Ni**	**Mo**
4118	0.20	0.80	0.22	0.50	. . .	0.11
5120	0.20	0.80	0.22	0.80	. . .	. . .
8620	0.20	0.80	0.22	0.50	0.55	0.20
4620	0.20	0.55	0.22	. . .	1.82	0.25

4320	0.20	0.55	0.22	0.50	1.82	0.25
3310	0.10	0.52	0.22	1.57	3.50	. . .
SCM420	0.20	0.72	0.25	1.05	. . .	0.22
20MnCr5	0.20	1.25	0.27	1.15	. . .	. . .

Source: Ref 2

Surface hardness on the order of 58 to 64 HRC is present for either high-carbon or carbonizing bearing steels; however, there are some differences in the core hardnesses of these two types of materials. The data in Table 3 show that the strength and toughness of the core are greater in the quenched and tempered low-carbon steels than in the unhardened portion of an induction-hardened high-carbon steel bearing part.

Table 3 Core properties of carburized versus induction-hardened components

Material	Hardness, HB	Yield strength		Ultimate tensile strength		Impact energy		Machining, %[a]
		MPa	**ksi**	**MPa**	**ksi**	**J**	**ft · lbf**	
8620[b]	30-45 HRC	825-965	120-140	1105-1240	160-180	55-110	40-80	65
5160[c]	197	275	40	725	105	10	7	55
1095[c]	192	380	55	655	95	3	2	45
52100[c]	197	345	50	635	95	. . .	. . .	40

Source: Ref 2

(a) Where 1212 carbon steel represents 100%.

(b) Quenched and tempered.

(c) Annealed.

From a surface fatigue standpoint, it is necessary to maintain sufficient hardness on and below the surface to resist abrasive/adhesive wear and to minimize contact fatigue in the whole region of the surface and below so that the material strength at each point is higher than the surface or subsurface shear stresses that occur from a loaded Hertzian surface stress. Vingsbo (Ref 1) points out that significant plastic deformation accompanies all forms of wear mechanisms; for contact fatigue it is evident in structural changes and small cell formation. Thus, strength as required can be considered to related to the yield strength of the metal, and the relationship suggested by Pederson and Rice (Ref 3) indicates how this strength varies with hardness (Fig. 1). Because of this, it is not necessarily the steel type that provides the given strength in an operating bearing but how it is treated. If the required hardness is provided from the surface to below the maximum subsurface stresses, then the fatigue resistance of the bearing steel will in general be sufficient.

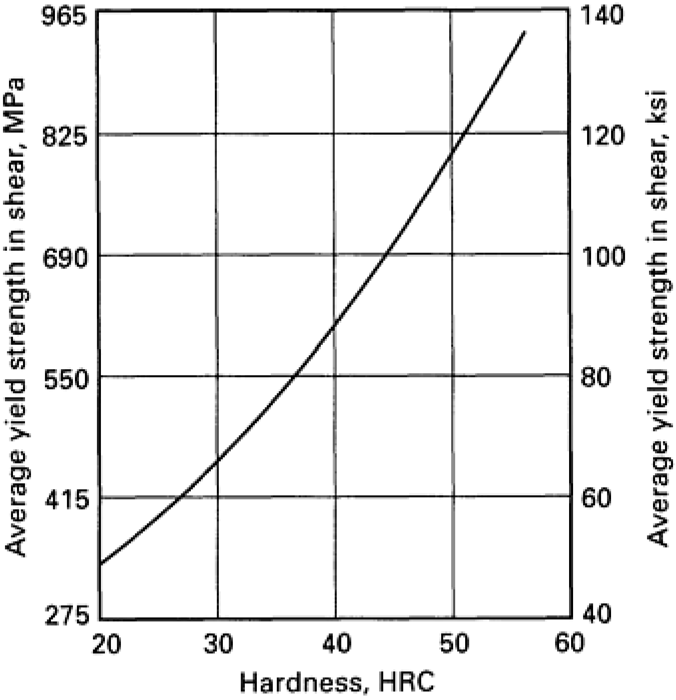


Fig. 1 Plot of shear yield strength versus hardness for typical bearing steel. Source: Ref 3

With proper heat treatment and temper, bearings made from the steels in Tables 1 and 2 should be able to operate in the temperature range of -50 °C (-60 °F) to ~120 to 150 °C (~250 to 300 °F). They should have minimum surface plastic deformation up to Hertzian contact stresses of 2 to 3 GPa (300 to 450 ksi). The Hertzian stress level has a profound effect on the contact fatigue of bearing steels, depending somewhat on bearing type; these differences will be addressed in the section "Concentrated Contacts" in this article.

Beyond the standard bearing steels of Tables 1 and 2, there have been specialty steels developed for either extended fatigue life or excessive operating conditions of temperature and corrosion. These are shown in Tables 4 and 5. In general, these steels have a much higher alloy content, in addition to their carbon content. The high-temperature steels are essentially from 9.3 to 13.5% alloy. CBS-600 is much lower at 4.2%, which is within the range of standard steels; BG42 is higher at 20.5% alloy. Along with different chemistry, these steels usually have significantly smaller and fewer inclusions. Because the bearings made of these materials are used in advanced aircraft engines or aerospace application with ultrahigh speeds, greatly extended fatigue life is required. Rather than air-melt quality, these steels tend to be, for example, vacuum arc remelted (VAR) or vacuum induction melted/vacuum arc remelted (VIM/VAR) for cleanness (that is, freedom from harmful nonmetallic inclusions). Methods of rating the cleanness of such steels are given in Ref 2. Because of the high speeds and the soak back temperatures from adjacent components, these steels usually require the higher temperature capability of about 315 to 370 °C (600 to 700 °F), as indicated in Table 4.

Table 4 Nominal compositions of high-temperature bearing steels

Steel	Composition, %								Maximum operating temperature[a]	
	C	Mn	Si	Cr	Ni	Mo	V	Other	°C	°F
M50	0.85	. . .	. . .	4.10	. . .	4.25	1.00	. . .	315	600
M50-NiL	0.13	0.25	0.20	4.20	3.40	4.25	1.20	. . .	315	600
Pyrowear 53	0.10	0.35	1.00	1.00	2.00	3.25	0.10	2.00 Cu	205	400
CBS-600	0.19	0.60	1.10	1.45	. . .	1.00	. . .	0.06 Al	230	450
Vasco X2-M	0.15	0.29	0.88	5.00	. . .	1.50	0.5	1.50 W	230	450
CBS-1000M	0.13	0.55	0.50	1.05	3.00	4.50	0.40	0.06 Al	315	600
BG42	1.15	0.50	0.30	14.5	. . .	4.00	1.20	. . .	370	700

Source: Ref 2

(a) Maximum service temperature, based on a minimum hot hardness of 58 HRC.

Table 5 Nominal compositions of corrosion-resistant bearing steels

Grade	Composition, %						
	C	**Mn**	**Si**	**Cr**	**Mo**	**V**	**Nb**
BG42	1.15	0.50	0.30	14.50	4.00	1.2	. . .
440C	1.00	0.40	0.30	17.00	0.50	. . .	. . .
440C modified	1.05	0.40	0.30	14.00	4.00	. . .	. . .
CRB-7	1.10	0.35	0.30	14.00	2.0	1.0	0.25

As noted in Table 5, all corrosion-resistant steels have high amounts of chromium. Their alloy content, excluding the high-carbon content, ranges from 17.9 to 20.5%. As might be expected, all these specialty steels with a significant increase in alloy content are more expensive than the standard bearing steels so that the special steels are used only in order to meet the severe requirements of extreme temperature and corrosion. When life requirements are extended, very often six to 10 times the standard duration, then steels with significantly lower levels of nonmetallic inclusions are needed. An example of the extension in fatigue life for cleaner steels is given in Fig. 2 (Ref 4).

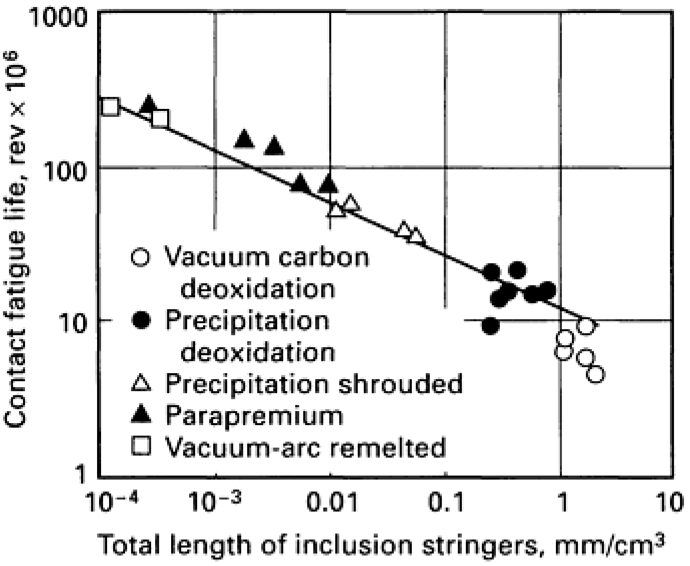


Fig. 2 Effect of total length of inclusion stringers on the rolling contact fatigue life of a 220 mm (8.7 in.) bore tapered roller bearing inner race as a function of steel cleanliness. Source: Ref 4

Cleanness level can be measured by the size and distribution of inclusion stringers that are left from the processing of steel. The presence of inclusions in a bearing component can be quantified by means of enhanced ultrasonic detection methods (Ref 4). Quantified ultrasonic results are supplemented with inclusion imaging and length verification using acoustic microscopy (see the article "Acoustic Microscopy" in this Volume). These tools, along with energy-dispersive or wavelength-dispersive chemical analyses using scanning electron microscopy (SEM), have confirmed that aluminum oxide stringers relate directly with bearing fatigue life under operating conditions that promote material fatigue. The size distribution and the total quantity of inclusion stringers in a given group of bearings has been shown to relate to performance (fatigue life) through a summation of the inclusion stringer lengths per unit volume of steel.

The total length in inclusion stringers per unit volume is the abscissa on Fig. 2. The steel cleanness range starts with an earlier vacuum carbon deoxidation steel process method of about 10 years ago to the precipitation-shrouded version of

today's air-melt steels and to today's vacuum-arc remelted steel. As shown in Fig. 2, the overall fatigue life range is more than an order of magnitude. VIM/VAR steel lives exceed even the VAR steel lives so that even though the present-day air-melt steels do give better lives, the vacuum-processed steels are meeting the fatigue life requirements for the extended-life applications.

Bearing steels are sometimes used for gears, cams, and other concentrated-contact applications. However, whenever the requirements of much higher surface sliding (tangential forces) are present in these other contacts, other slightly different chemistries may be specified and certainly differing steel treatments and processing will be required to better meet the contact conditions of sliding, rougher surfaces, and even the presence of bending fatigue. In general, bearing steels are limited to rolling contact bearings based on surface hardness requirements, harden-ability, and cleanness level in order to minimize nonmetallic inclusion fatigue.

Concentrated Contacts

In an operating rolling contact bearing, the instantaneous action takes place within the concentrated contacts formed between the bearing raceway and roller or ball surfaces. Figure 3 is a simplified diagram of such a contact showing two surfaces with curvature brought together by load to form a Hertzian stress field with maximum contact pressure, *P*, as shown. The longer length of contact is usually transverse to the direction of motion, and the narrower width of contact, 2*b*, for smaller to medium size bearings, will be on the order of 0.1 to 1 mm (0.004 to 0.04 in.). The contact tends to be a rectangle for line contact and an ellipse for point contact (usually ball bearings). However, if there is a curvature or profile on one or both surfaces in the transverse direction, even roller bearings can have essentially elliptical or truncated elliptical contact, depending on the load level.

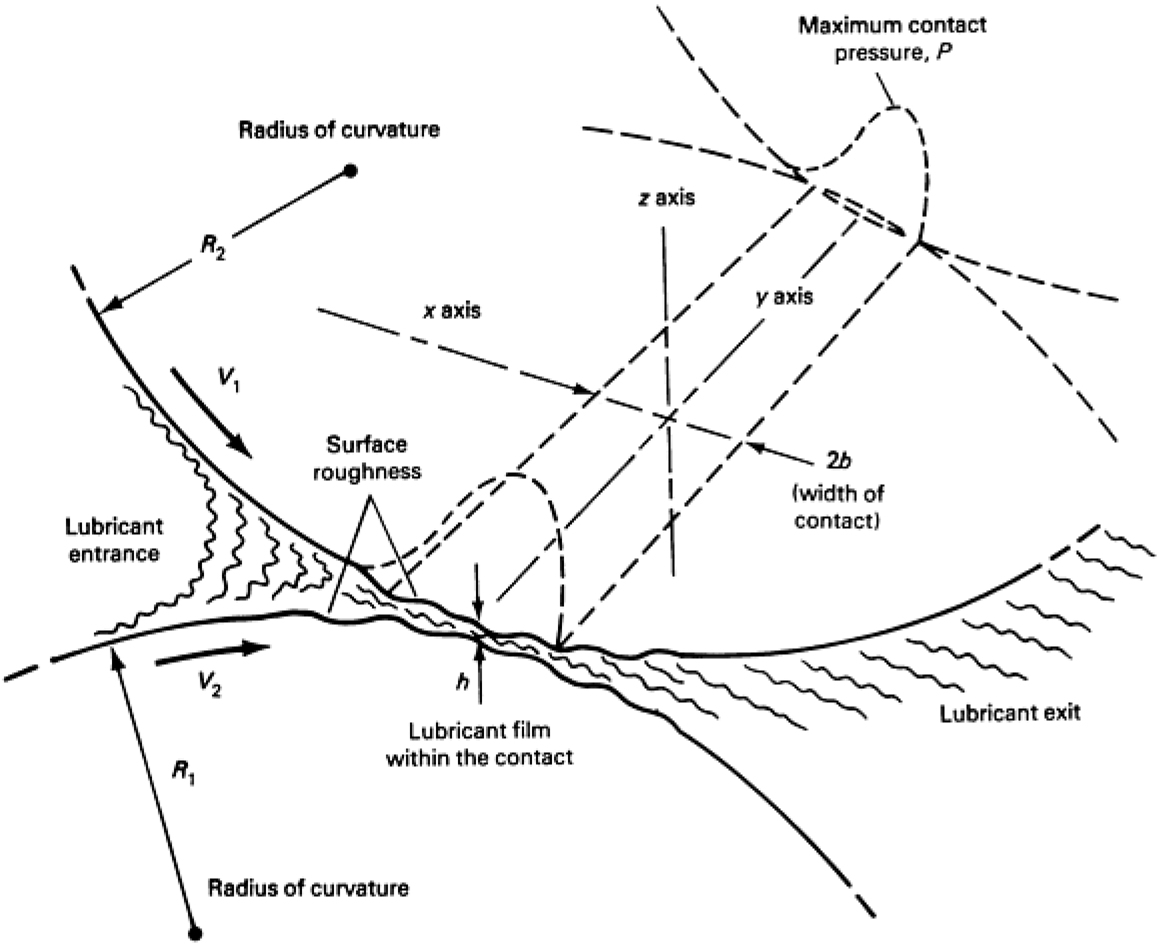


Fig. 3 Schematic showing key components of tribological contact between a raceway and the rolling elements of a steel bearing

The moving surfaces entrain lubricant into the contact and form a lubricant film with thickness, *h,* that partially or completely separates the surfaces from contacting. The film thickness can be calculated using elastohydrodynamic theory (Ref 5, 6, 7) that recognizes the exponential increase in the lubricant viscosity within the Hertzian field and the local elastic deformation of the bearing steel surfaces that form the contact area. (Details of how to determine the thickness of a lubricant film are given in Ref 5, 6, 7.) The contacting steel surfaces in a bearing are finished by grinding, which is usually followed by some superfinishing process (for example, honing). Larger bearings may only have a ground surface. Whatever the finishing method, there is some roughness left on the bearing steel surfaces, as indicated in Fig. 3. The peaks or asperities of the surface may penetrate the lubricant film, depending on the surface roughness and the thickness of the film controlled by the operating viscosity; the pressure viscosity at operating temperature; the surface velocities, V_1 and V_2; the radii of curvature; and to a lesser extent, the prevailing load.

The lubricant film thickness divided by the combined roughness of the two surfaces has been called the specific film thickness or the lambda ratio, Λ. Specific film thickness has been a popular method for describing the operating conditions within the contact for over 20 years (Ref 8). Numerous wear and fatigue tests have been reported comparing the result for $0.1 \leq \Lambda < 20$. Figure 4 is an example of eight test groups of a specific size ball bearing (45 × 100 mm, or 1.8 × 3.9 in., deep-groove ball bearing of through-hardened 52100 steel) (Ref 9), and Fig. 5 shows 11 test groups of a specific roller bearing size (46 × 85 mm, or 1.8 × 3.3 in., tapered roller bearing of carburized and hardened 8620 steel) (Ref 10). For these specific bearings, the fit of the test groups was by confidence bands (0.95) for the ball bearings and a regression coefficient (0.88) for the tapered roller bearings, indicating that for each of these bearings a reasonable predicting equation was developed:

- For the ball bearing data:

$$L_{10} = C \cdot \Lambda^{0.54} \qquad \textbf{(Eq 1)}$$

- where L_{10} is the bearing life for 90% reliability, Λ is the lambda ratio ranging from ~0.7 to 20, and C is the constant developed from the specific tests
- For the tapered roller bearing data:

$$L_R = C \cdot \Lambda^{0.86} \qquad \textbf{(Eq 2)}$$

- where L_R is the bearing life in terms of ratio to catalog expected life (that is, relative fatigue life), Λ is the lambda ratio ranging from 0.04 to 0.3, and C is the constant developed from the specific tests

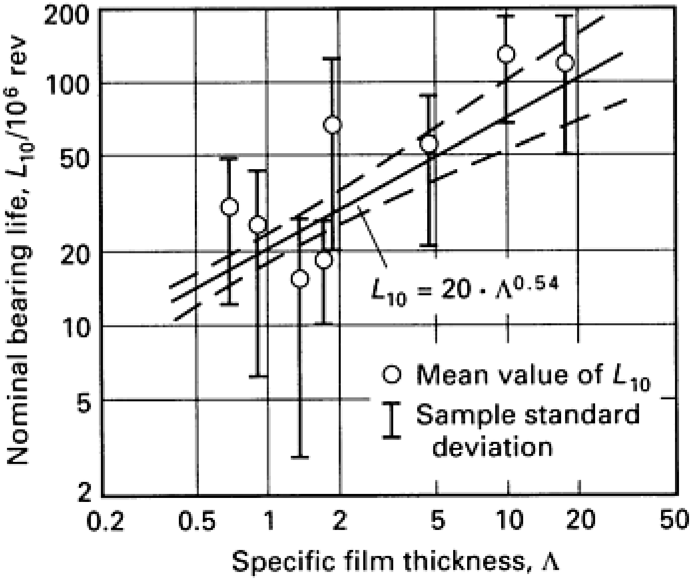


Fig. 4 Plot of normal bearing life versus specific film thickness for 45 × 100 mm ($1\frac{3}{4}$ × 3.9 in.) deep-groove ball bearings made of through-hardened 52100 bearing steel. Data were compiled for eight test groups totaling 6309 bearing units. Confidence limits are 95%. Source: Ref 9

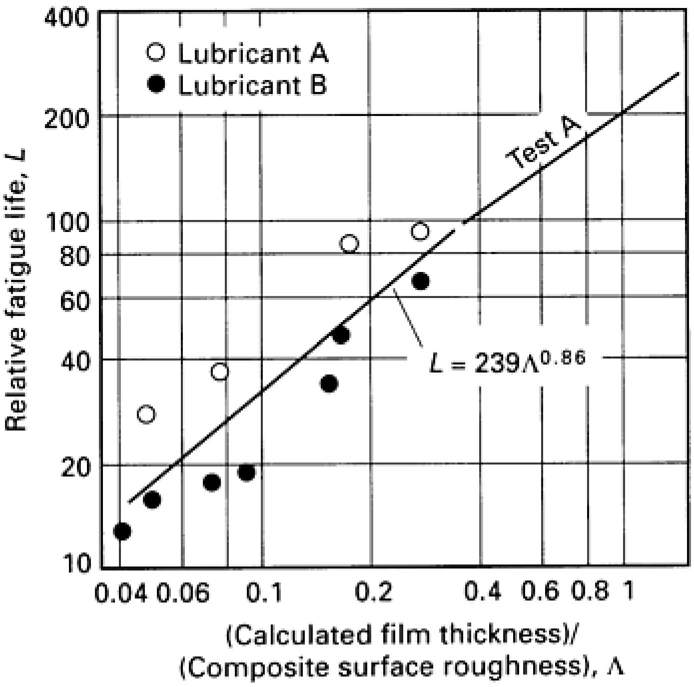


Fig. 5 Plot of relative fatigue life versus specific film thickness, Λ, for 46 × 85 mm (1.8 × 3.3 in.) tapered roller bearings made of carburized and hardened 8620 steel. Data were compiled for 11 test groups with 88% correlation coefficient on the regression line. Test A represents results from previous testing. Source: Ref 10

Because Λ represents the operating condition within a bearing steel contact and because the simple lambda ratio seems to relate well to fatigue life (as Fig. 4 and 5 indicate), there have been attempts to combine much published data and develop a general relationship between Λ and fatigue wear. Figure 6 shows a significant attempt to combine the results from seven different sources of both ball bearing and roller bearing tests (Ref 11). Note that Λ does relate to the relative life factor, $N(\Lambda)/N_{cat}$, but there is considerable scatter due to other factors that cause variation in fatigue life. There is no question, however, that bearing steel contact fatigue wear is significantly affected by the contact lubricant operating conditions.

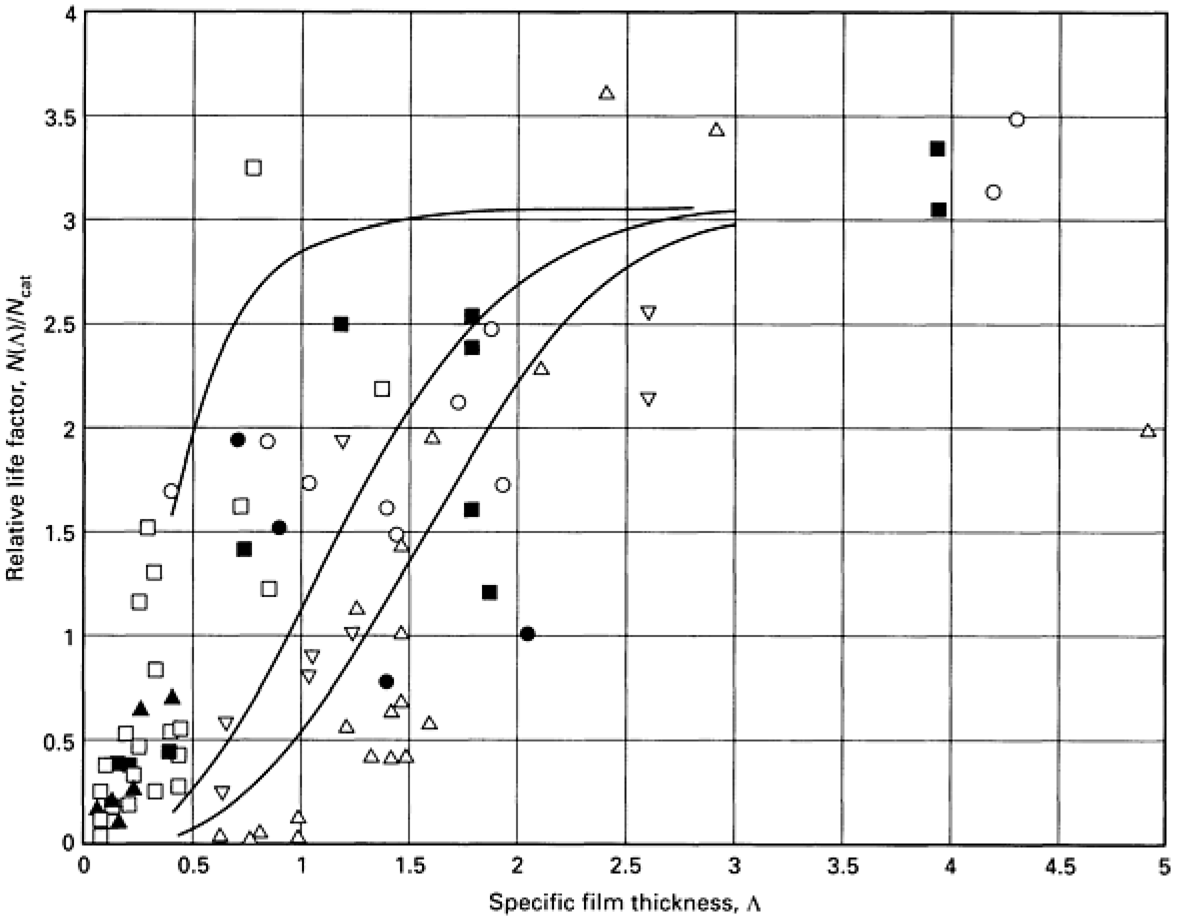


Fig. 6 Plot of relative lives versus specific film thickness as a function of selected asperity slope curves: A, 48', the smoothest surface; B, 2°, the median surface; C, 3°30', the roughest surface. Data points shown represent work of seven researchers. Source: Ref 11

Roughness or surface topography is not an intrinsic property of the material or surface (Ref 12) but depends on how it was processed and finished and on how the range of wavelengths of the topography interact with the bearing steel contact, especially the width 2*b* in the rolling direction (Ref 13). Tallian (Ref 11) has proposed the use of asperity slope to more completely describe roughness as it influences Λ(Fig. 6). The inference is that bearing steel surfaces having sharper asperities may have lower life than surfaces having broader asperities. Although these curves did not agree with some of the actual groups of test bearings, the premise is sound. This scatter also indicates that Λmay need to be modified to better represent roughness for the range of bearing sizes and contacts that are represented in Fig. 6 (Ref 8).

Lambda Ratio and Modes of Wear

Another important factor relating to Λconcerns the change in mode of contact fatigue for different values of Λ. This also relates to the presence of the other bearing steel wear mechanisms that depend on the value of Λ. Table 6 is a simplified listing of the modes of wear or damage classification for bearings. When $\Lambda$is > 3, almost complete separation of the surfaces occurs and minimal adhesive or abrasive wear takes place. Bearing steel contact fatigue wear is essentially all subsurface at the regions of maximum shear so that all fatigue is material related and primarily inclusion initiated. This subsurface region is roughly 0.25 to 0.5 times the width of contact (2*b*) below the surface. Under these conditions, the steel is expected to meet or exceed the rating life for the bearings. The experimentally determined load-life equation below is valid for current bearings of steel described in Tables 1 and 2:

$$L_{10} = \left(\frac{C}{P}\right)^n \cdot 10^6 \text{ rev} \qquad \textbf{(Eq 3)}$$

where L_{10} is the rating life with 90% reliability (in revolutions), C is the basic dynamic capacity, C_R (radial) or C_T (thrust), based on the rating equations given in Ref 14 and 15, P is the equivalent radial load or equivalent thrust load to match C_R or C_T, and n is the load exponent (experimentally based with consensus values published in the bearing standards) (Ref 14 and 15); typically, n = 3 for ball bearings and n = 10/3 for roller bearings.

Table 6 Wear mode or damage classification for bearing steels

Contract fatigue	
Subsurface origin: Inclusion	Within macroshear classical shear stress zone below contact surface
	Near surface zone of microshear greatly affected by surface roughness (asperites)
Subsurface origin: Subcase fatigue	Origin near case-core interface if yeld strength is exceeded by applied stress
Surface origin	Origin at the end of contact aggravated by edge geometry
	Multiple origins of micropitting (peeling or frosting)
	Point surface origin at localized stress risers (dents, grooves, surface inclusions)
Plastic flow	
Brinelling or debris denting	
Load excursions above the material yield point	
Yielding aggravated by high-temperature excusions	
Wear	
Adhesive	Normal: Mild or "controlled" (usually identified as "run-in")
	Severe: Irreversible scuffing, scoring, smearing, or seizure
Abrasive	Normal: Ususally three-body system, mediaum to fine particles that are also associated with "run-in"
	Severe: Grooving, gouging, denting with ridges that cause serious surface stress risers
Corrosive	Water or acidic constitutents from lubricant breakdown or temperature, chemically aggressive additives in lubricant
Fretting	Microscale adhesive and abrasice wear
	Corrosion involvement depending on environment and contacts

Considering the exponents on the load ratio, it is clear that load is the most significant factor in bearing life (contact fatigue). From Eq 3, for loads of 0.1 to 0.4 C, the life ratios in Table 7 hold.

Table 7 Effect of load on bearing life

Life ratio, C/P	Ball bearings $(C/P)^3$	Roller bearings $(C/P)^{10/3}$
10.0	1000	2150
2.5	16	21

Thus, for ball bearings, reducing the load by one-fourth will results in >60 times bearing life and for roller bearings, one-fourth less load results in >100 times bearing life. The data in Table 7 assume only inclusion, bearing steel related fatigue, and contact fatigue wear as represented by Eq 3.

With $\Lambda < 3$, more asperity interaction will take place in the contact (per Fig. 3 and Table 6; inclusions in the near-surface zone may also be involved). With $\Lambda < 1.0$, there can be significant changes to the surface traction so that the level and region of subsurface and near-surface shear stress increases, leading to reduction in fatigue life. The change in one particular stress, orthogonal shear reversal range, is indicated on Fig. 7. As the coefficient of friction, μ, reflecting the increased tangential surface force goes up, the surface shear stress could eventually match that of the subsurface so that a much larger volume of material below the bearing steel surface is subjected to higher shear stresses. For example, under normal operating conditions, the coefficient of friction between the raceway and roller for tapered roller bearings has been measured from 0.001 to 0.005 (Ref 17). The bearings were 76 mm (3.0 in.) bore by 162 mm (6.38 in.) outside diameter in size. The increase in the coefficient of friction followed speed changes from 1500 to 500 rev/min, the lambda ratio, Λ, decreasing from 2.5 to 0.5. These measured values are low and probably include other friction losses besides the surface traction. However, the decrease in the value of Λ was clearly accompanied by greater surface traction and increased asperity interaction. An indication of how such a change in traction with Λ held constant might reduce bearing life is shown in Fig. 8. The data points are from Ref 18 and show measured traction coefficients for tests with slide/roll ratios that varied from 0 to 0.8%. The three points show actual relative fatigue lives $N/N_{\mu e=0}$ for rolling balls tested at 6 GPa

(870 ksi). The approximate Λ over the slide/roll ratio range was estimated at 1.25 (Ref 11). Therefore, surface friction forces can be changed by the amount of sliding, by Λ changes, or usually by the combination of both.

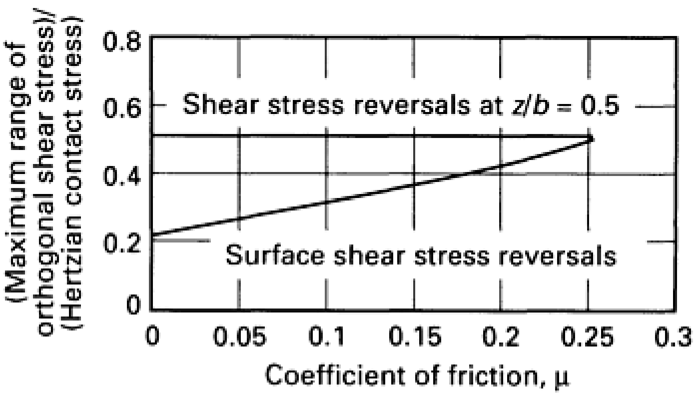


Fig. 7 Plot of the coefficient of friction versus the ratio of the maximum range of orthogonal shear stress to the Hertzian contact stress. (*z*/*b*) = 0.5 is the distance below the surface in the *z* direction (Fig. 3) relative to contact width 2*b*. Source: Ref 16

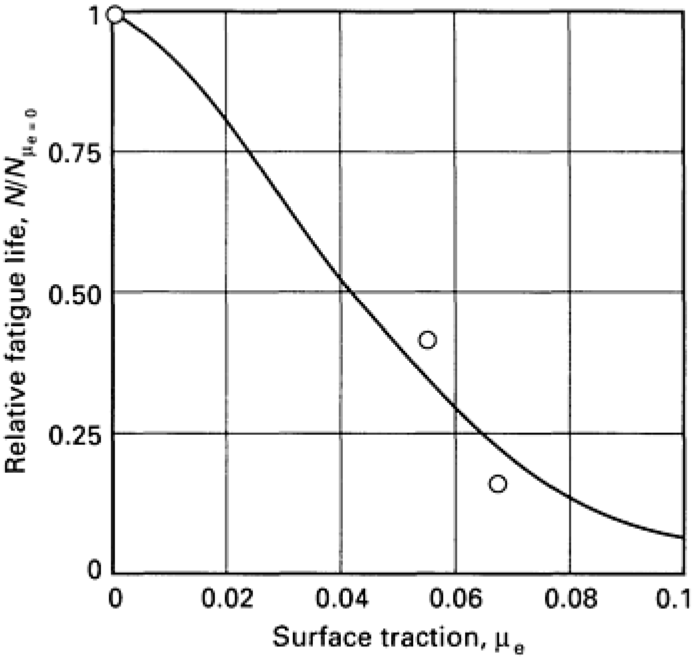


Fig. 8 Plot of relative fatigue life, $N/N_{\mu e=0}$, versus surface traction. $N_{\mu e=0}$ is fatigue life when traction equals zero. Model parameters: asperity slope, 2°; specific film thickness, Λ, 1.25. Source: Ref 11, 18

For $\Lambda < 1.0$, because of the increased asperity contact, and if significant increases in traction can occur, subsurface fatigue gives way to near-surface fatigue (high stress pockets below the highest, sharpest asperities). With Λ even lower, failure can change to surface fatigue, often initiated by surface defects. Dents, deep grinding grooves, handling nicks, surface inclusions, and so on, are now the prime stress raisers rather than nonmetallic inclusions within the steel. The failure modes resulting from the surface defects are described in Table 6. One mode is point surface origin fatigue in which the fatigue propagation fans out from the point in an arrowhead fashion. Other surface-related fatigue can come from misalignment in a ball or roller bearing that reaches the end of the raceway (geometric stress point); or because of the topography (roughness of the asperities), fatigue can develop as multiple pitting, which appears as frosting or peeling.

Under these conditions, fatigue life drops as indicated in Fig. 5 and 6. Equation 3 is no longer valid and L_{10} must be adjusted because rating life will no longer be attained. The bearing industry recognized the need to be able to adjust L_{10} when the concept of life adjustment factors was introduced in 1972 (Ref 19). Equation 4 indicates the general approach to life adjustment:

$$L_{na} = a_1 \cdot a_2 \cdot a_3 \cdot \left(\frac{C}{P}\right)^n \cdot 10^6 \text{ rev} \quad \textbf{(Eq 4)}$$

Or by adjusting rating life, L_{10}:

$$L_{na} = a_1 \cdot a_2 \cdot a_3 \cdot L_{10} \quad \textbf{(Eq 5)}$$

where L_{na} is the calculated life (in rev) adjusted by the multiplicative factors, a_1, a_2, and a_3; a_1 is the reliability factor for other than 90% reliability; a_2 is the material factor for other than standard bearing steels; and a_3 is the application or environmental factor, primarily recognizing the lubricant condition within a bearing; a_3 also can include alignment, internal load distribution changes, temperature, contaminants (liquid and solid), and fatigue spall criteria.

Reliability Factor a_1. The reliability factor is based on the life scatter experienced by a group of apparently identical bearings running under the same load, same Λ or same environmental conditions. This scatter is quantified by the Weibull distribution, specifically the Weibull slope (m) or dispersion parameter. The general equation for this is:

$$a_1 = [9.4912 \ln(100/R)]^{1/m} \quad \textbf{(Eq 6)}$$

where a_1 is the reliability factor to adjust the L_{10} life, R is the reliability expressed in terms of percent survival, and m is the Weibull slope, for reliability >90%. The value of m is 1.5 in the bearing standards. Background data for establishing the Weibull slopes for ball bearings and roller bearings are in Ref 20 and 21, respectively.

Material Factor a_2. Recognizing that cleaner steel (that is, steel with fewer and smaller inclusions) is being used in today's bearings, most bearing manufacturers have increased ratings in the last several years so that the fatigue constants present in various rating equations reflect the increased fatigue life expected from the present steels. The 1990 rating standards reflect this in a generic way within the present bearing ratings given in Ref 14 and 15. Because of the considerable experimental basis required to establish specific bearing ratings (that is, the fatigue constant, the Weibull slope, and the load-life exponent, along with specific material heat treatment and processing differences), the determination of contact fatigue wear of bearing steel alone is difficult to extract from the rating equations. Thus, a_2 is not the kind of factor that can be used as a typical wear rate coefficient. The material factor a_2 was initially developed to recognize the ultraclean steels (that is, VAR, VIM/VAR, or electroslag remelting, or ESR) that could give significantly longer lives or increased reliability in special bearing applications that could justify the increased cost of such steels. This is still the primary use of a_2 in the aerospace industry and other industries requiring extended-life bearings.

Application of Environmental Condition Factor a_3. A simple summary of the subfactors that make up the factor a_3 is given in Table 8. The subfactors listed relate directly with the bearing contacts. As would be expected, the lubricant--usually the elastohydrodynamic lubricant (EHL) film compared to the surface roughness, very often as Λ--is an important subfactor. However, the lubricant can be a mineral oil, synthetic, or other fluid, with properties of viscosity, pressure viscosity, surface wetting ability, surface tension, and shear strain-stress relationships that are enhanced or degraded depending on the additive mix, their particular formulation, and how any of these change over time. Unfortunately, there is only limited information about how changes in these properties impact Λ or relate to bearing steel fatigue or wear.

Table 8 Operating conditions impacting bearing performance that are subfactors of application or environmental factor, a_3

Lubricant--surface topography	**These control how surfaces interact (EHL film/composite roughness as lambda ratio is the primary model used)**
Temperature--speed	Bulk conditions such as these control the concentrated contacts of a bearing
Load zone--alignment	Factors impacting load and actual stress within all the active contacts in a bearing
Fatigue spall criteria	Definitions of initiation and propagation as these impact the actual lives of bearings in an application
Contaminants	Liquid and solid that can flow through a contact

Bearings operating at a given speed (and load) tend to reach an operating bulk temperature for the bearing steel that establishes the temperature for the lubricant coming into the contact. Although this temperature may be difficult to determine, any attempts at calculating the lubricant film within the bearing require as accurate an assessment of this operating temperature as possible. The heat generated within a bearing can be better understood by the following equation:

$$Q_{gen} = \text{function}(S)^{3/2}, (\mu)^{1/2}, (F)^{1/3} \quad \textbf{(Eq 7)}$$

where Q_{gen} is the heat generation (in watts), S is the rotational speed, μ is the lubricant viscosity at bearing operating temperature (in centipoise), and F is the load (in Newtons).

The point of this simple function is to recognize how bearing steel temperature will change depending on changes to speed, viscosity, and load and that the contribution of these factors to temperature change is not the same. Therefore, changes in any of the factors given in Eq 7 will affect operating temperature, the lubricant film, and the operating conditions represented by Λ.

Depending on the external loading on a bearing, radial or thrust, with overturning moments, along with any internal clearance (or lack thereof), there can be sizeable internal load variation, shifting the amount of loading on each rolling element. Depending on the magnitude of the overturning moments or deflections, excessive misalignment can occur. Table 9 gives approximate allowable misalignment for bearings before reduction in life can occur. This misalignment results in nonuniform stress being applied to the steel, and fatigue or wear occurs at the region of the highest stresses.

Table 9 Allowable misalignment in rolling-element bearings before bearing life is affected

Bearing type	Misalignment		
	Radians	**in./in.**	**Degrees**
Cylindrical and tapered bearings	0.0005-0.001	0.0005-0.001	0.03-0.06
Ball bearings	0.003-0.005	0.003-0.005	0.17-0.29

Fatigue spall criteria are normally defined as the first evidence of fatigue or by the size of the spall, such as 3 mm ($\frac{1}{8}$ in.) diameter (Ref 9) or 6 mm^2 (0.01 in.^{2}) (Ref 10). Whether a bearing has any useful life beyond this beginning spall size depends on the operating conditions that control the rate of fatigue spall propagation. For mild operating conditions with lower temperature, reasonable speeds, and lubricant film that provide $\Lambda > 1.0$, it may be possible to run until 10% or some portion of a raceway is spalled. However, under thin-film conditions, especially at higher temperatures and heavy loads ($C/P < 4.0$), fatigue propagation may be so rapid that no extension of bearing life beyond the first detection of a fatigue spall can be expected.

Contaminants within a steel bearing can cause considerable damage and lower fatigue life. Rough running, excessive friction, overheating, and other forms of wear can occur from contaminants; therefore, proper filtration and sealing are important. The extent of filtration required depends on the bearing size and type selected. Acid formation from lubricant breakdown and the presence of water can be especially damaging. The steel can experience surface etching and corrosive wear. An indication of the impact of water alone is shown in Fig. 9.

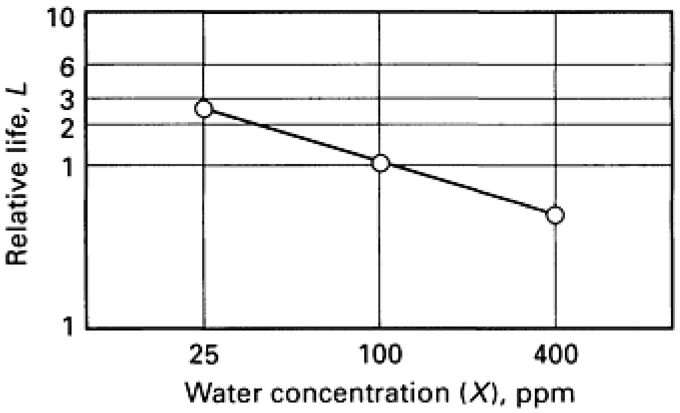


Fig. 9 Plot of relative bearing life, *L*, versus water concentration, *X*, where $L = (100/X)^{0.6}$. At 100 ppm concentration, $L = 1$. Source: Ref 22

Solid contaminants or debris are also damaging in their ability to collect in the inlet region of a bearing contact, thereby causing lubricant starvation (Ref 23). Debris may also cause dents on the bearing steel surface that are large enough to be stress raisers, lowering contact life significantly. Debris that passes continuously through a bearing contact tends to be shattered or deformed into much smaller pieces. These form an exponential-type distribution within a closed or circulating lubricant system. The debris particle sizes can range from >100 μm (>0.004 in.) (usually only a few particles approach this size) to a middle range of ~50 to 5 μm (~2000 to 200 μin.) (from hundreds to thousands of particles) to particles <5 μm (<200 μin.) (depending on the source, quantities can be into the millions). The largest particles cause the dents that affect fatigue life. The middle range can contribute to abrasive wear in bearings (and fatigue loss from denting in bearings with smaller contacts, especially small ball bearings). Finally, the smallest size particles cause mild wear in larger bearings and serious abrasive wear in small bearings with contact lengths <5 mm (<0.2 in.).

The ability to determine the loss of fatigue life from debris denting is in the developing stage. A subfactor of a_3 that can account for the size and side slope of the dent relative to the contact size and the tolerance of the material (both in terms of hardness and fracture toughness) will be required (Ref 24, 25, 26). Carburized bearing steels with considerably higher fracture toughness (Fig. 10) tend to be more tolerant than through-hardened steels to the consequences of debris denting. This tolerance is also dependent on the amount of retained austenite present in the surface layer.

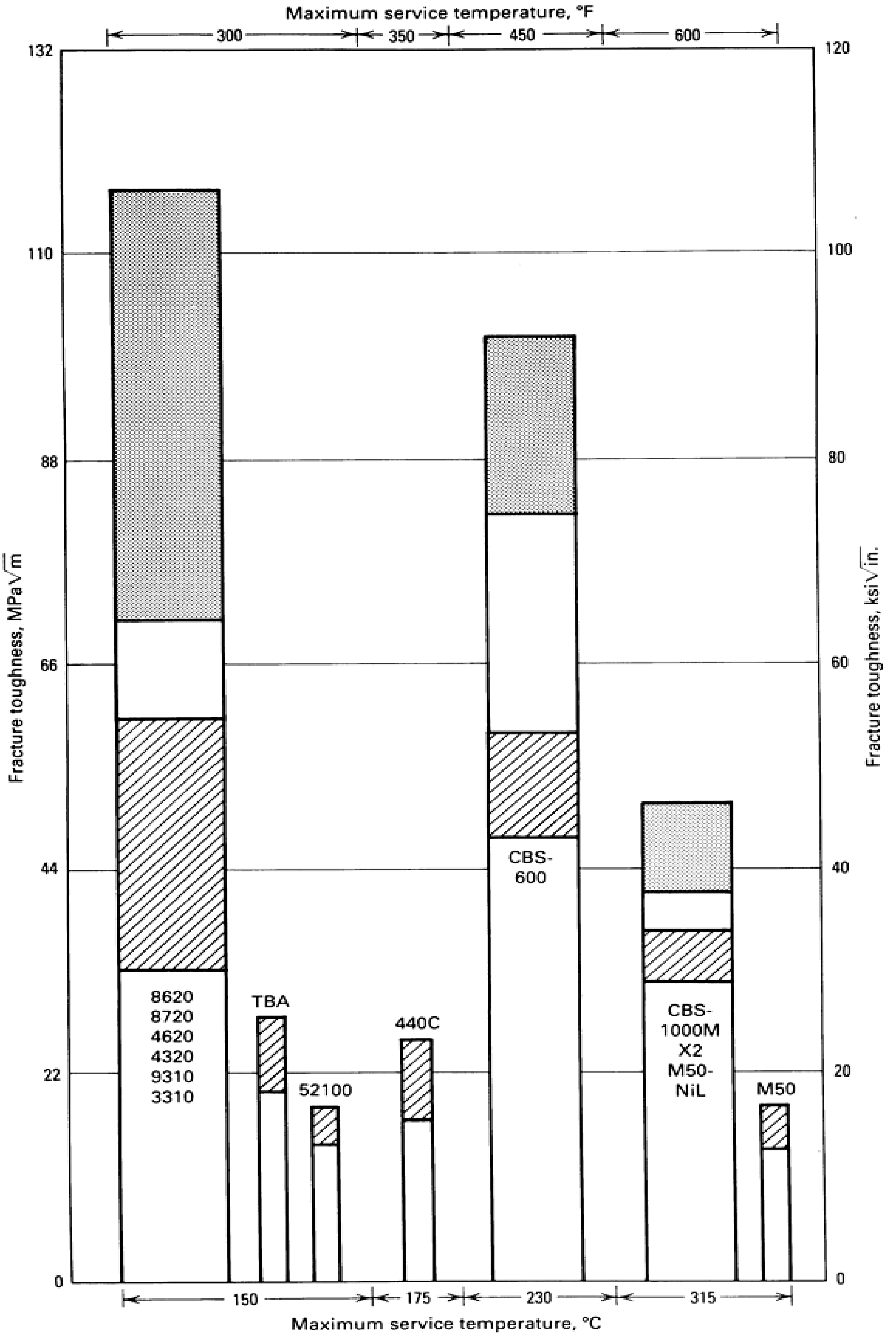


Fig. 10 Composite fracture toughness of carburizing and homogeneous high-carbon steels in slow bending. Case depth is 0.76 to 0.89 mm (0.030 to 0.035 in.) to 0.50% C level. Shaded areas indicate range of *K* values for cracks originating in core. Cross-hatched areas indicate range of *K* values for cracks originating in case. Charpy-sized specimens were carburized, hardened, tempered, and precracked to several depths in case and

core regions before testing. As cracks progress inward, the fracture resistance of carburized composites improves significantly.

Abrasive Wear

In most applications, rolling bearings do not experience appreciable abrasive or adhesive wear. There are some wear particles formed from the run-in that bearings undergo in the initial period of operation. If there is some denting on the bearing, wear can also take place. A 45 × 100 mm (1.8 × 4 in.) ball bearing and a 45.2 × 73 mm (1.78 × 2.9 in.) tapered roller bearing were each dented by a Vickers hardness indenter at two places on the inner raceway. The ball and tapered roller bearings were then each run in a clean, closed lubricant system (Ref 27), and they developed wear particles that in total weighed about 15 and 25 mg (0.23 and 0.39 gr), respectively (Ref 24). It was estimated that the ball bearing generated 2×10^5 particles, in a size range of 10 to 30 μm (400 to 1200 μin.), running with a Λ value of 1.8; the tapered roller bearing generated about 8×10^4 particles in a size range about 5 to 50 μm (200 to 2000 μin.), running with a Λ value of ~1.1 (Ref 27). Therefore, with reasonable lambda ratio and no debris added, bearings should experience only mild wear during run-in or with some denting. If debris of sufficient size and quantity enters a lubricant system, then a bearing will have considerable wear.

A series of bearing wear tests have been conducted in a two-chamber bearing wear machine of the type shown in Fig. 11 (Ref 28). Each chamber has a small 33.3 × 68.3 mm (1.31 × 2.69 in.) tapered roller bearing with 250 cm^3 (8.5 oz) of an SAE 80 EP gear oil (1.3×10^{-5} m^2/s, or 13 cSt) in the closed system. The lubricant contains a measured amount of debris, usually silica dust, of known size range. The debris is kept in suspension by the flingers attached to the integral rotating shaft. The test runs at 2700 rev/min with $\Lambda \sim 0.8$, a thrust load giving 1 GPa (150 ksi) stress, and for 40 h (6.5×10^6 rev duration). Four bearings tested with silica dust in a sequence of 2, 5, 10, and 40 h, which cleaning and weighing at each stop, show the high rate of initial wear and the gradual leveling off by 40 h (Fig. 12). There has not been a change in the bearing steel, but the debris particles are so crushed that most pass through the contact film and can no longer cause wear.

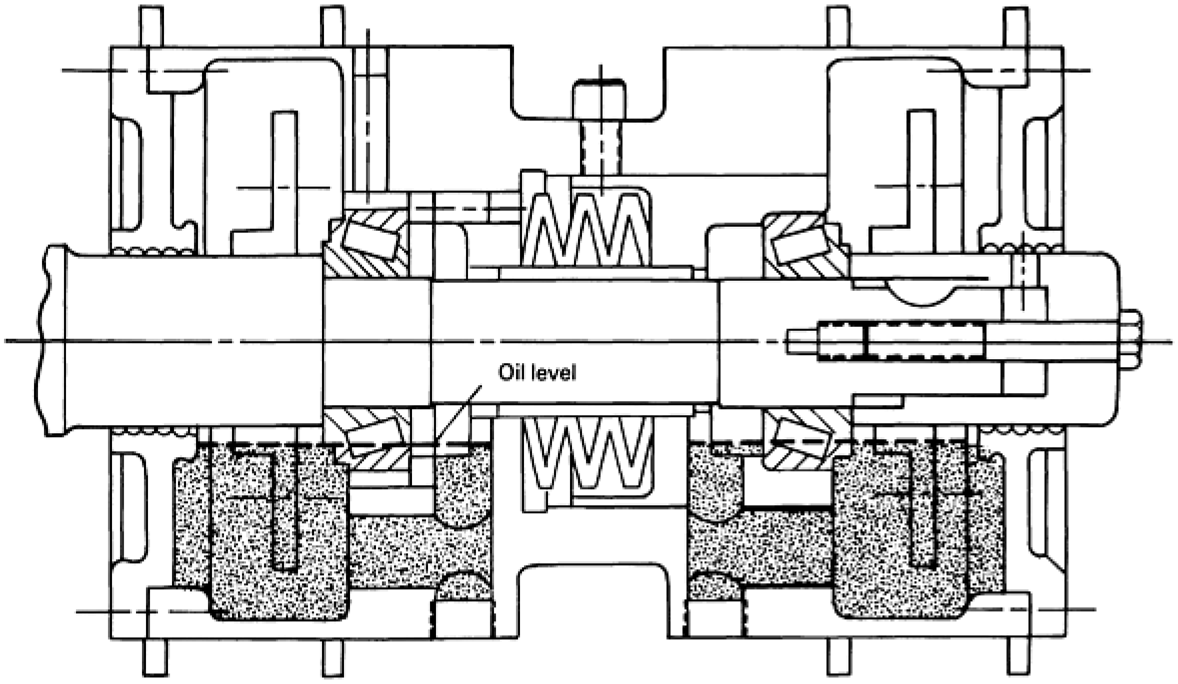


Fig. 11 Schematic showing cross-sectional view of a two-chamber bearing wear machine

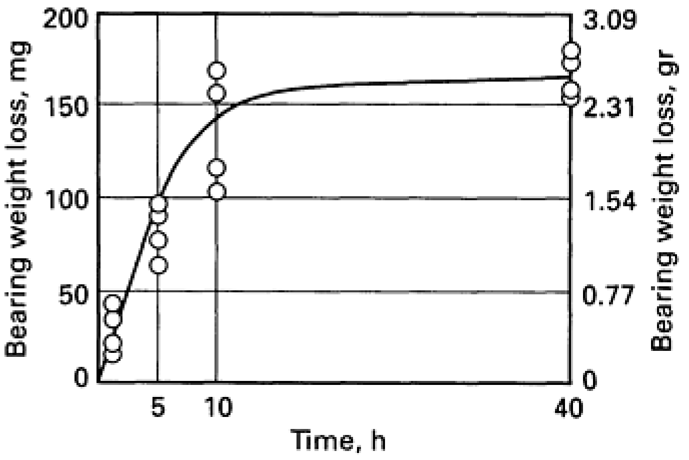


Fig. 12 Plot of bearing wear versus time for 33.3 × 68.3 mm (1.31 × 2.69 in.) tapered roller bearings using data obtained with the bearing wear machine shown in Fig. 11. Test conditions: lubricant, SAE 80 EP gear oil with 1.3 × 10^{-5} m^2/s (13 cSt) viscosity used at temperatures of 90 to 100 °C (190 to 210 °F); contaminant, uncut silica dust with concentration of 1000 mg/L (60 gr/gal)

Repeat tests demonstrated that reasonable wear data could be generated with this tester. Tests with no debris and with debris concentrations of 250, 500, 750, and 1000 mg/L (15, 30, 45, and 60 gr/gal) indicated that a linear relationship between debris (particle) concentration and wear (milligrams of weight loss) exists with a regression correlation coefficient of 0.95. Additional tests of the same debris concentration range, but with different cuts of silica dust (particle size ranges) indicated the same linear relationship but no significant difference between the debris size ranges (Fig. 13).

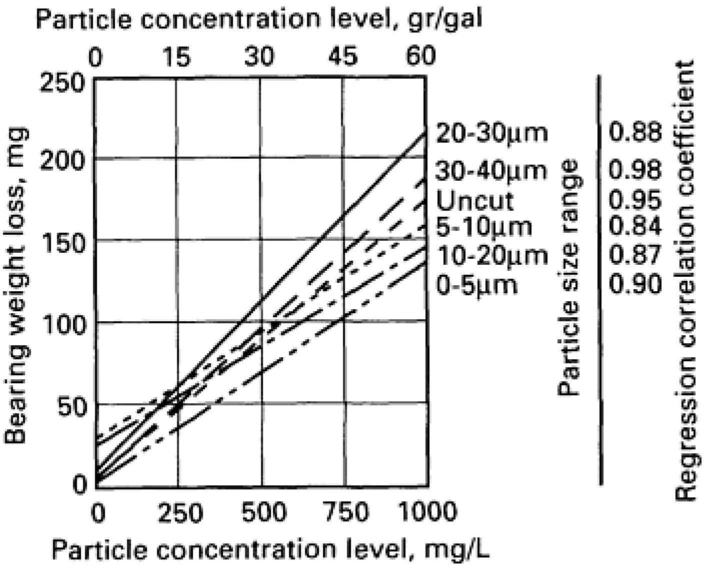


Fig. 13 Plot of bearing weight loss versus particle concentration level as a function of particle size range. Test parameters: lubricant, SAE 80 EP gear oil used at temperature of 90 to 100 °C (190 to 210 °F); contaminant, silica dust; test duration, 40 h

The maximum weight loss of about 200 mg (3 gr) for this small bearing would cause borderline internal dimensional changes in the bearing. the debris concentration, 1000 mg/L (60 gr/gal), if considered from a volume standpoint, may be a better comparison. The volume of silica to the volume (liter) of lubricant would be about 450 ppm. However, weight losses of two to three times this 200 mg (3 gr) level could cause serious internal geometry damage to the bearing.

Figure 14 indicates the amount of wear that occurs with debris contaminants of different hardness level. The silica fine uncut dust line can be considered the reference line. The bearing steel debris then caused about 60% of the wear from silica, and the carbonyl iron produced no more than 5% of silica wear. The diamond dust caused extreme wear. The bearings were badly damaged and the tests were stopped considerably before the 10 h time limit. Figure 15 shows how different lubricants can affect abrasive wear of bearing steel. The amount of wear increases as viscosity decreases for the mineral oils and the extreme-pressure (EP) gear oil. The differences are shown in Table 10.

Table 10 Effect of selected lubricants on the abrasive wear of bearing steels

Lubricant	Viscosity at 95 °C (200 °F)		Wear[a]	
	SUS[d]	**Kinetic viscosity, cSt**	**mg**	**gr[e]**
SAE 140[b]	180	38.3	60	0.90
SAE 50[c]	110	22.7	75	1.16
SAE 30[c]	71	13.2	95	1.46
SAE 5[c]	41	4.5	125	1.93
SAE 90[b]	85	16.8	145	2.23
SAE 80[b]	65	11.6	170	2.62

(a) Bearing weight loss at 1000 mg/L (60 gr/gal) debris concentration.

(b) EP gear oil.

(c) Mineral oil.

(d) SUS, Saybolt Universal seconds.

(e) gr, grain

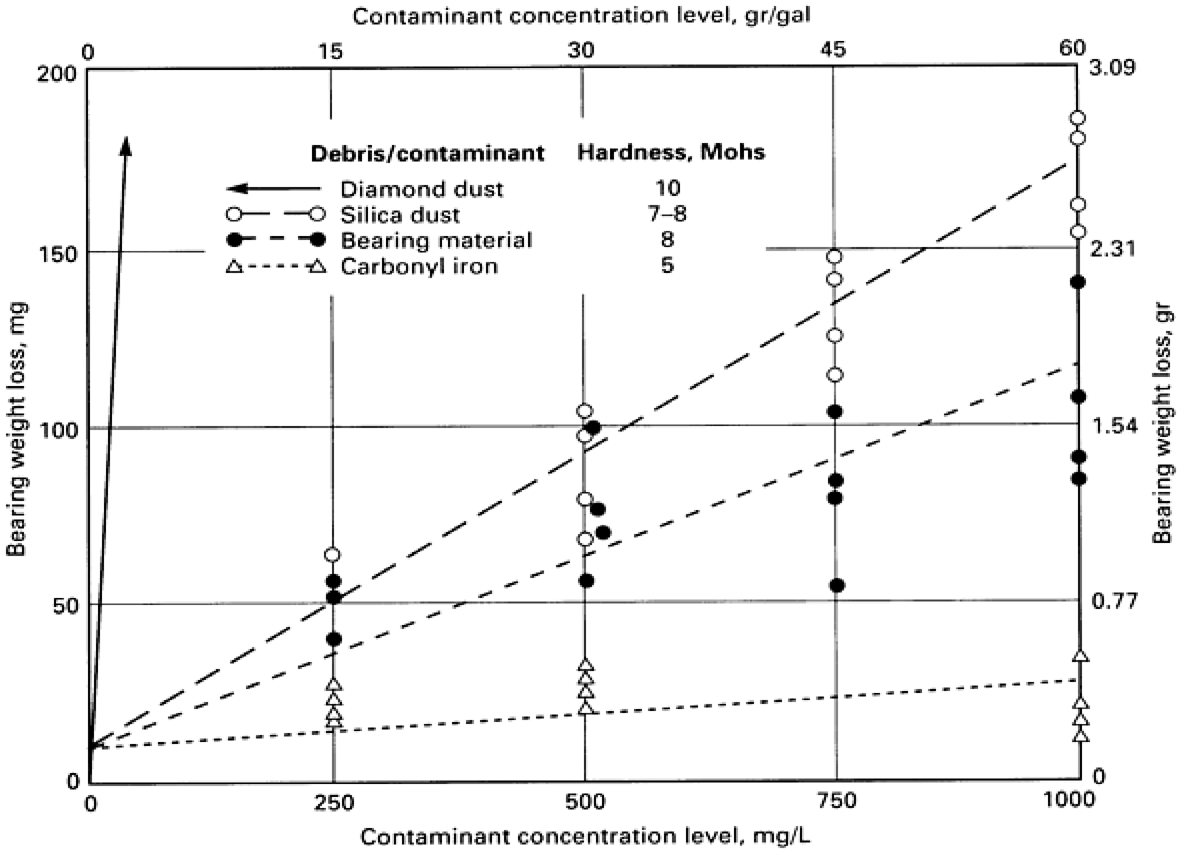


Fig. 14 Plot of bearing weight loss versus contaminant/debris concentration level as a function of four selected contaminants of varying hardness. Test parameters: lubricant, SAE 80 EP gear oil used at temperature of 90 to 100 °C (190 to 210 °F); test duration, 10 h

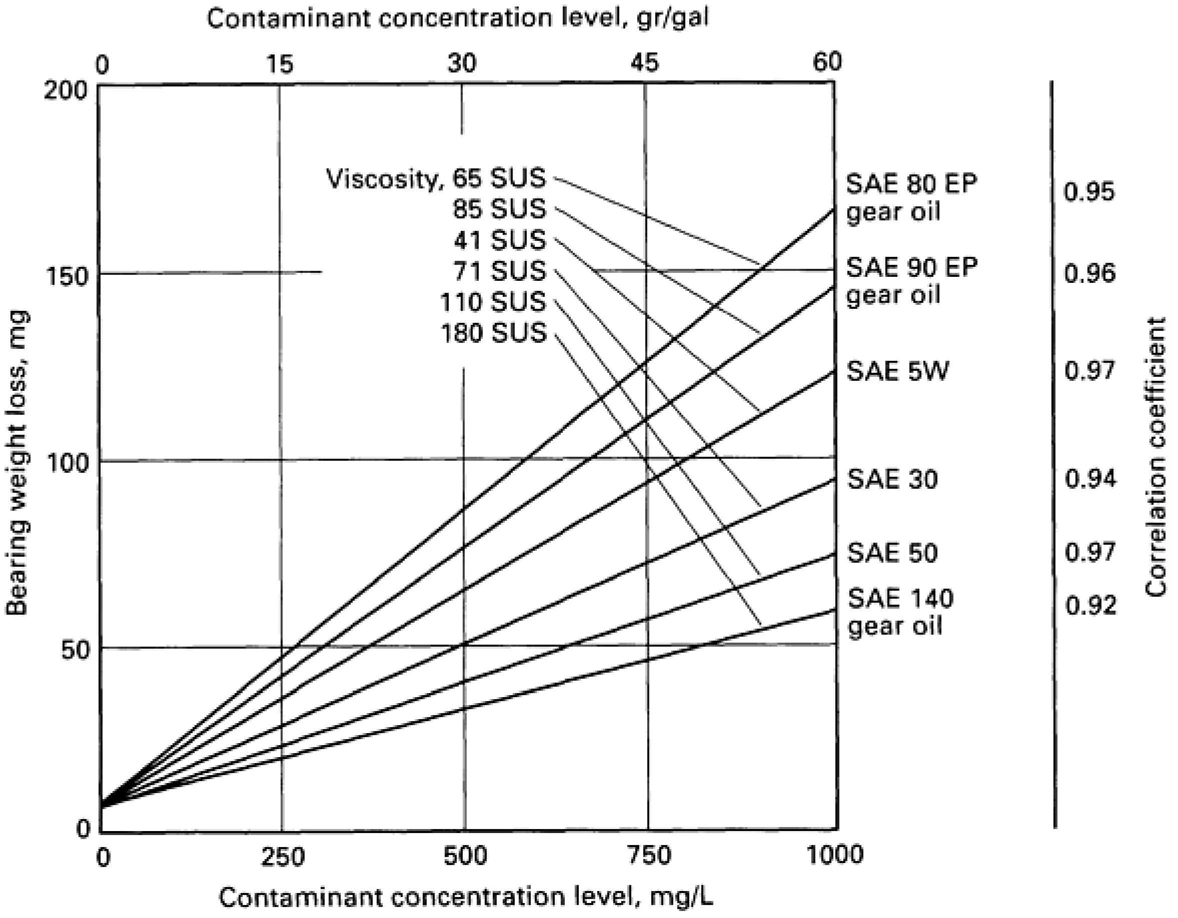


Fig. 15 Plot of bearing weight loss versus contaminant concentration level as a function of selected lubricant oils of varying viscocity. Test parameters: lubricant temperature of 90 to 100 °C (190 to 210 °F); contaminant, uncut silica dust; test duration, 10 h. Oil viscosity data are at temperature of 95 °C (200 °F)

There is also some indication from Table 10 that the extreme-pressure additive in an oil may lead to increased wear in the presence of debris. Abrasive wear can remove the oxide films normally present on bearing steel surfaces, and the chemical activity of the additive constituents (for example, sulfur) will attack the exposed surface and promote the increased wear.

Adhesive Wear

Under certain circumstances, bearing surfaces show evidence of adhesive wear. It may be observed under low lambda ratios, very often <0.3, combined with lower inadequate viscosity and higher temperatures.

Low-Speed Wear. Adhesive wear can occur under extremely low-speed conditions when adequate lubricant films cannot be formed and asperity interaction (probably multiple occurrences within a contact) leads to junction formation, deformation, and fracture. The adhesive wear manifests itself by the metal transfer, which, if small (down to microscopic in size) and somewhat isolated, may not grow but be limited as run-in. If it grows, the result is usually rough operation and erratic torque. If cages or rolling-element retainers are involved, bearing lock-up can finally occur.

High-Speed Wear. At the other extreme, under high-speed conditions, especially with centrifugal force influence, intermittent rolling-element lift-off can occur and skidding or extended sliding develops, which can lead to excessive heat generation and finally seizure. Standard bearing steels (Tables 1 and 2) are more susceptible to adhesive wear than the high-temperature steels because the high-temperature steels maintain hardness and strength to better resist the formation of localized microweld junctions within the contact. It is also possible that, at temperatures which exist at the microweld junctions, if certain EP additives are present in the lubricants, their activation may resist adhesive wear in bearing steel.

Some indication of the ranges within which adhesive wear may not occur is shown in a scuffing limit map such as that proposed in Ref 29.

References

1. O. Vingsbo, Wear and Wear Mechanisms, *Proceedings Wear of Materials Conference,* K. Ludema, R. Glaeser, and R. Rhee, Ed., American Society of Mechanical Engineers, 1979
2. H.I. Burrier, Jr., Bearing Steels, *Metals Handbook,* Vol 1, 10th ed., *Properties and Selection,* 1990, p 380-388
3. R. Pederson and S.L. Rice, Case Crushing of Carburized and Hardened Gears, *SAE Trans.,* Vol 68, 1960, p 187-220
4. J.D. Stover, R.V. Kolarik, and D.M. Keener, "The Detection of Aluminum Oxide Stringers in Steel Using an Ultrasonic Measuring Method, "Preprint 31st Mechanical Working and Steel Processing Conference, American Institute of Mining, Metallurgical, and Petroleum Engineers
5. B.J. Hamrock and D. Dowson, Isothermal Elastohydrodynamic Lubrication of Point Contacts, Part I--Theoretical Formulation, *J. Lubr. Technol. (Trans. ASME),* Vol 98 (No. 2), Apr 1976, p 223-229
6. B.J. Hamrock and D. Dowson, Isothermal Elastohydrodynamic Lubrication of Point Contacts, Part III--Fully Flooded Results, *J. Lubr. Technol. (Trans ASME),* Vol 99 (No. 2), Apr 1977, p 264-276
7. D. Dowson and S. Toyoda, A Central Film Thickness Formula for Elastohydrodynamic Line Contacts, *Proceedings of 5th Leeds-Lyon Symposium,* 1978, p 60-65
8. C.A. Moyer and L.L. Bahney, Modifying the Lambda Ratio to Functional Line Contacts, *STLE Trib. Trans.,* Vol 33 (No. 4) 1990, p 535-542
9. S. Andreason and T. Lund, Ball Bearing Endurance Testing Considering Elastohydrodynamic Lubrication, Paper C36 *IMechE EHD Symposium Proceedings* 1972, p 138-141
10. C.A. Danner, Fatigue Life of Tapered Roller Bearings under Minimal Lubricant Films, *ASLE Trans.,* Vol 13 (No. 4), Oct 1970, p 241-251
11. T.E. Tallian, Rolling Bearing Life Modifying Factors for Film Thickness, Surface Roughness and Friction, *J. Lubr. Technol. (Trans. ASME),* Vol 103, Oct 1981, p 509-520
12. R.S. Sayles and T.R. Thomas, Surface Topography as a Nonstationary Random Process, *Nature,* Vol 271, Feb 1978, p 431-434
13. R.H. Leaver, R.S. Sayles, and T.R. Thomas, Mixed Lubrication and Surface Topography of Rolling Contacts, *IMechE Proc.,* Vol 188, 1974, p 461-469
14. "Load Ratings and Fatigue Life for Ball Bearings," ANSI/AFBMA Std 9-1990, American National Standard, 17 July 1990
15. "Load Ratings and Fatigue Life for Roller Bearings," ANSI/AFBMA Std 11-1990, American National Standard, 17 July 1990
16. J.W. Kannel and J.L. Tevaarwerk, Subsurface Stress Evaluations under Rolling/Sliding Conditions, *J. Lubr. Technol. (Trans. ASME),* Vol 106, Jan 1984, p 96-103
17. F.J. Wren and C.A. Moyer, Understanding Friction and EHL. Films in Concentrated Contacts of a Tapered Roller Bearing, *1972 EHL Symposium* (Leeds, England), Institute of Mechanical Engineers, 11 Apr 1972, p 55-67
18. E.N. Diaconescu, G.D. Kerrison, and P.B. Macpherson, A New Machine for Studying the Effects of Sliding and Traction on the Fatigue Life of Point Contacts with Initial Test Results, *Trans. ASLE,* Vol 18 (No. 4), 1975, p 239-248
19. "Load Ratings and Fatigue Life for Roller Bearings," AFBMA Standard 11, June 1972
20. T. Tallian, Weibull Distributions of Rolling Contact Fatigue Life and Deviations Therefrom, *ASLE Trans.,* No. 5, 1962, p 183-196
21. C.A. Moyer, The Status and Future of Roller Bearing Life Prediction, Proceedings of the International Industrial Tribology Symposium at Northwestern University (Aug 1990), STLE Special Publication SP-31, *Advances in Engineering Tribology,* Y. Chung and H.S. Cheng, Ed., Apr 1991, p 89-99

22. R.E. Cantley, The Effect of Water in Lubricating Oil on Bearing Fatigue Life, *STLE Trans.,* Vol 20 (No. 3), July 1977, p 244-248
23. G.T.Y. Wan and H.A. Spikes, The Behavior of Suspended Solid Particles in Rolling and Sliding Elastohydrodynamic Contacts, *STLE Tribol. Trans.,* Vol 31 (No. 1), Jan 1988, p 12-21
24. C.A. Moyer, "The Influence of Debris on Rolling Bearing Performance: Identifying the Relevant Factors," Paper 871687, Off-Highway/Power Plant Congress (Milwaukee), Society of Automotive Engineers, 14-17 Sept 1987
25. C.A. Moyer, "The Influence of Debris on Roller Bearing Performance: Chapter II," Paper 891906, International Off-Highway/Power Plant Congress (Milwaukee), Society of Automotive Engineers, 11-14 Sept 1989
26. A.A. Lubrecht, C.H. Venner, S. Lane, B. Jacobsen, and E. Ioannides, Surface Damage-Composition of Theoretical and Experimental Endurance Lives of Rolling Bearings, *Proceedings Japan International Tribology Conference,* 1990, p 185-190
27. H. Dalal and P. Senholzi, Characteristics of Wear Particles Generated during Failure Progression of Rolling Bearings, *ASLE Trans.,* Vol 20 (No. 3), 1977, p 233-243
28. B. Fitzsimmons and H.D. Clevenger, Contaminated Lubricants and Tapered Roller Bearing Wear, *ASLE Trans.,* Vol 20 (No. 2), 1977, p 97-107
29. W. Hirst and A.E. Hollander, Surface Finish and Damage in Sliding, *Proc. R. Soc. (London) A,* Vol A337, 1974, p 379-394

Friction and Wear of Tool Steels

James L. Maloney III, Latrobe Steel Company

Introduction

TOOLS STEELS are carbon, alloy, and high-speed steels that can be hardened and tempered to high hardness and strength values. These steels are used when resistance to deformation, heat, and wear are necessary. Typical applications involve metalworking operations, such as cutting, shearing, and forming. However, tool steels are also used in other situations, such as pump impellers and cases where good resistance to wear is necessary.

Most tool steels are electric-furnace melted, refined in a secondary vessel, cast into ingots, and hot worked into final form. However, some are used as-cast and others are manufactured by:

- Atomizing the steel into powder with either gas or water
- Consolidating by sintering, pressing, or forging
- Finishing by hot working
- Heat treating (if in near-net shape)

All tool steels are heat treated after the tool is rough-machined by hardening (austenitizing and quenching into water, oil, or air, as appropriate) and tempering to obtain the requisite hardness and toughness for the tooling application.

The American Iron and Steel Institute (AISI) has developed a classification system for tool steels (Ref 1). The commonly used tool steels in each classification are shown in Table 1, whereas the relative abrasion resistance of various tool steels is shown in Fig. 1. Wear-related applications, in order of ascending abrasion resistance based on the AISI classification, are described below.

Table 1 Composition limits of principal types of tool steels

Designations			Composition[a], %								
AISI	**SAE**	**UNS**	**C**	**Mn**	**Si**	**Cr**	**Ni**	**Mo**	**W**	**V**	**Co**
Molybdenum high-speed steels											
M1	M1	T11301	0.78-0.88	0.15-0.40	0.20-0.50	3.50-4.00	0.30 max	8.20-9.20	1.40-2.10	1.00-1.35	. . .
M2	M2	T11302	0.78-0.88; 0.95-1.05	0.15-0.40	0.20-0.45	3.75-4.50	0.30 max	4.50-5.50	5.50-6.75	1.75-2.20	. . .
M3, class 1	M3	T11313	1.00-1.10	0.15-0.40	0.20-0.45	3.75-4.50	0.30 max	4.75-6.50	5.00-6.75	2.25-2.75	. . .
M3, class 2	M3	T11323	1.15-1.25	0.15-0.40	0.20-0.45	3.75-4.50	0.30 max	4.75-6.50	5.00-6.75	2.75-3.75	. . .
M4	M4	T11304	1.25-1.40	0.15-0.40	0.20-0.45	3.75-4.75	0.30 max	4.25-5.50	5.25-6.50	3.75-4.50	. . .
M6	. . .	T11306	0.75-0.85	0.15-0.40	0.20-0.45	3.75-4.50	0.30 max	4.50-5.50	3.75-4.75	1.30-1.70	11.00-13.00
M7	. . .	T11307	0.97-1.05	0.15-0.40	0.20-0.55	3.50-4.00	0.30 max	8.20-9.20	1.40-2.10	1.75-2.25	. . .
M10	. . .	T11310	0.84-0.94; 0.95-1.05	0.10-0.40	0.20-0.45	3.75-4.50	0.30 max	7.75-8.50	. . .	1.80-2.20	. . .
M30	. . .	T11330	0.75-0.85	0.15-0.40	0.20-0.45	3.50-4.25	0.30 max	7.75-9.00	1.30-2.30	1.00-1.40	4.50-5.50
M33	. . .	T11333	0.85-0.92	0.15-0.40	0.15-0.50	3.50-4.00	0.30 max	9.00-10.00	1.30-2.10	1.00-1.35	7.75-8.75
M34	. . .	T11334	0.85-0.92	0.15-0.40	0.20-0.45	3.50-4.00	0.30 max	7.75-9.20	1.40-2.10	1.90-2.30	7.75-8.75
M36	. . .	T11336	0.80-0.90	0.15-0.40	0.20-0.45	3.75-4.50	0.30 max	4.50-5.50	5.50-5.60	1.75-2.25	7.75-8.75
M41	. . .	T11341	1.05-1.15	0.20-0.60	0.15-0.50	3.75-4.50	0.30 max	3.25-4.25	6.25-7.00	1.75-2.25	4.75-5.75
M42	. . .	T11342	1.05-1.15	0.15-0.40	0.15-0.65	3.50-4.25	0.30 max	9.00-10.00	1.15-1.85	0.95-1.35	7.75-8.75
M43	. . .	T11343	1.15-1.25	0.20-0.40	0.15-0.65	3.50-4.25	0.30 max	7.50-8.50	2.25-3.00	1.50-1.75	7.75-8.75
M44	. . .	T11344	1.10-1.20	0.20-0.40	0.30-0.55	4.00-4.75	0.30 max	6.00-7.00	5.00-5.75	1.85-2.20	11.00-12.25
M46	. . .	T11346	1.22-1.30	0.20-0.40	0.40-0.65	3.70-4.20	0.30 max	8.00-8.50	1.90-2.20	3.00-3.30	7.80-8.80
M47	. . .	T11347	1.05-1.15	0.15-0.40	0.20-0.45	3.50-4.00	0.30 max	9.25-10.00	1.30-1.80	1.15-1.35	4.75-5.25
Tungsten high-speed steels											
T1	T1	T12001	0.65-0.80	0.10-0.40	0.20-0.40	3.75-4.00	0.30 max	. . .	17.25-18.75	0.90-1.30	. . .
T2	T2	T12002	0.80-0.90	0.20-0.40	0.20-0.40	3.75-4.50	0.30 max	1.00 max	17.50-19.00	1.80-2.40	. . .
T4	T4	T12004	0.70-0.80	0.10-0.40	0.20-0.40	3.75-4.50	0.30 max	0.40-1.00	17.50-19.00	0.80-1.20	4.25-5.75
T5	T5	T12005	0.75-0.85	0.20-0.40	0.20-0.40	3.75-5.00	0.30 max	0.50-1.25	17.50-19.00	1.80-2.40	7.00-9.50
T6		T12006	0.75-0.85	0.20-0.40	0.20-0.40	4.00-4.75	0.30 max	0.40-1.00	18.50-21.00	1.50-2.10	11.00-13.00
T8	T8	T12008	0.75-0.85	0.20-0.40	0.20-0.40	3.75-4.50	0.30 max	0.40-1.00	13.25-14.75	1.80-2.40	4.25-5.75
T15	. . .	T12015	1.50-1.60	0.15-0.40	0.15-0.40	3.75-5.00	0.30 max	1.00 max	11.75-13.00	4.50-5.25	4.75-5.25
Chromium hot-work steels											
H10	. . .	T20810	0.35-0.45	0.25-0.70	0.80-1.20	3.00-3.75	0.30 max	2.00-3.00	. . .	0.25-0.75	. . .
H11	H11	T20811	0.33-0.43	0.20-0.50	0.80-1.20	4.75-5.50	0.30 max	1.10-1.60	. . .	0.30-0.60	. . .
H12	H12	T20812	0.30-0.40	0.20-0.50	0.80-1.20	4.75-5.50	0.30 max	1.25-1.75	1.00-7.00	0.50 max	. . .
H13	H13	T20813	0.32-0.45	0.20-	0.80-	4.75-	0.30	1.10-	. . .	0.80-	. . .

				0.50	1.20	5.50	max	1.75		1.20	
H14		T20814	0.35-0.45	0.20-0.50	0.80-1.20	4.75-5.50	0.30 max	. . .	4.00-5.25	. . .	. . .
H19		T20819	0.32-0.45	0.20-0.50	0.20-0.50	4.00-4.75	0.30 max	0.30-0.55	3.75-4.50	1.75-2.20	4.00-4.50
Tungsten hot-work steels											
H21	H21	T20821	0.26-0.36	0.15-0.40	0.15-0.50	3.00-3.75	0.30 max	. . .	8.50-10.00	0.30-0.60	. . .
H22	. . .	T20822	0.30-0.40	0.15-0.40	0.15-0.40	1.75-3.75	0.30 max	. . .	10.00-11.75	0.25-0.50	. . .
H23	. . .	T20823	0.25-0.35	0.15-0.40	0.15-0.60	11.00-12.75	0.30 max	. . .	11.00-12.75	0.75-1.25	. . .
H24	. . .	T20824	0.42-0.53	0.15-0.40	0.15-0.40	2.50-3.50	0.30 max	. . .	14.00-16.00	0.40-0.60	. . .
H25	. . .	T20825	0.22-0.32	0.15-0.40	0.15-0.40	3.75-4.50	0.30 max	. . .	14.00-16.00	0.40-0.60	. . .
H26	. . .	T20826	0.45-0.55[b]	0.15-0.40	0.15-0.40	3.75-4.50	0.30 max	. . .	17.25-19.00	0.75-1.25	. . .
Molybdenum hot-work steels											
H42	. . .	T20842	0.55-0.70[b]	0.15-0.40	. . .	3.75-4.50	0.30 max	4.50-5.50	5.50-6.75	1.75-2.20	. . .
Air-hardening medium-alloy cold-work steels											
A2	A2	T30102	0.95-1.05	1.00 max	0.50 max	4.75-5.50	0.30 max	0.90-1.40	. . .	0.15-0.50	. . .
A3	. . .	T30103	1.20-1.30	0.40-0.60	0.50 max	4.75-5.50	0.30 max	0.90-1.40	. . .	0.80-1.40	. . .
A4	. . .	T30104	0.95-1.05	1.80-2.20	0.50 max	0.90-2.20	0.30 max	0.90-1.40	. . .	. . .	. . .
A6	. . .	T30106	0.65-0.75	1.80-2.50	0.50 max	0.90-1.20	0.30 max	0.90-1.40	. . .	. . .	. . .
A7	. . .	T30107	2.00-2.85	0.80 max	0.50 max	5.00-5.75	0.30 max	0.90-1.40	0.50-1.50	3.90-5.15	. . .
A8	. . .	T30108	0.50-0.60	0.50 max	0.75-1.10	4.75-5.50	0.30 max	1.15-1.65	1.00-1.50	. . .	. . .
A9	. . .	T30109	0.45-0.55	0.50 max	0.95-1.15	4.75-5.50	1.25-1.75	1.30-1.80	. . .	0.80-1.40	. . .
A10	. . .	T30110	1.25-1.50[c]	1.60-2.10	1.00-1.50	. . .	1.55-2.05	1.25-1.75	. . .	. . .	. . .
High-carbon, high-chromium cold-work steels											
D2	D2	T30402	1.40-1.60	0.60 max	0.60 max	11.00-13.00	0.30 max	0.70-1.20	. . .	1.10 max	1.00 max
D3	D3	T30403	2.00-2.35	0.60 max	0.60 max	11.00-13.50	0.30 max	. . .	1.00 max	1.00 max	. . .
D4	. . .	T30404	2.05-2.40	0.60 max	0.60 max	11.00-13.00	0.30 max	0.70-1.20	. . .	1.00 max	. . .
D5	D5	T30405	1.40-1.60	0.60 max	0.60 max	11.00-13.00	0.30 max	0.70-1.20	. . .	1.00 max	2.50-3.50
D7	D7	T30407	2.15-2.50	0.60 max	0.60 max	11.50-13.50	0.30 max	0.70-1.20	. . .	3.80-4.40	. . .
Oil-hardening cold-work steels											
O1	O1	T31501	0.85-1.00	1.00-1.40	0.50 max	0.40-0.60	0.30 max	. . .	0.40-0.60	0.30 max	. . .
O2	O2	T31502	0.85-0.95	1.40-1.80	0.50 max	0.35 max	0.30 max	0.30 max	. . .	0.30 max	. . .
O6	O6	T31506	1.25-1.55[c]	0.30-1.10	0.55-1.50	0.30 max	0.30 max	0.20-0.30	. . .	. . .	. . .
O7	. . .	T31507	1.10-1.30	1.00 max	0.60 max	0.35-0.85	0.30 max	0.30 max	1.00-2.00	0.40 max	. . .
Shock-resisting steels											
S1	S1	T41901	0.40-0.55	0.10-0.40	0.15-1.20	1.00-1.80	0.30 max	0.50 max	1.50-3.00	0.15-0.30	. . .
S2	S2	T41902	0.40-0.55	0.30-0.50	0.90-1.20	. . .	0.30 max	0.30-0.60	. . .	0.50 max	. . .
S5	S5	T41905	0.50-0.65	0.60-1.00	1.75-2.25	0.35 max	. . .	0.20-1.35	. . .	0.35 max	. . .
S6		T41906	0.40-0.50	1.20-	2.00-	1.20-	. . .	0.30-	. . .	0.20-	. . .

				1.50	2.50	1.50		0.50		0.40	
S7		T41907	0.45-0.55	0.20-0.80	0.20-1.00	3.00-3.50	. . .	1.30-1.80	. . .	0.20-0.30[(d)]	. . .
Low-alloy special-purpose tool steels											
L2	. . .	T61202	0.45-1.00[(b)]	0.10-0.90	0.50 max	0.70-1.20	. . .	0.25 max	. . .	0.10-0.30	. . .
L6	L6	T61206	0.65-0.75	0.25-0.80	0.50 max	0.60-1.20	1.25-2.00	0.50 max	. . .	0.20-0.30[(d)]	. . .
Low-carbon mold steels											
P2	. . .	T51602	0.10 max	0.10-0.40	0.10-0.40	0.75-1.25	0.10-0.50	0.15-0.40	. . .	. . .	. . .
P3	. . .	T51603	0.10 max	0.20-0.60	0.40 max	0.40-0.75	1.00-1.50	. . .	. . .	. . .	. . .
P4	. . .	T51604	0.12 max	0.20-0.60	0.10-0.40	4.00-5.25	. . .	0.40-1.00	. . .	. . .	. . .
P5	. . .	T51605	0.10 max	0.20-0.60	0.40 max	2.00-2.50	0.35 max	. . .	. . .	. . .	. . .
P6	. . .	T51606	0.05-0.15	0.35-0.70	0.10-0.40	1.25-1.75	3.25-3.75	. . .	. . .	. . .	. . .
P20	. . .	T51620	0.28-0.40	0.60-1.00	0.20-0.80	1.40-2.00	. . .	0.30-0.55	. . .	. . .	. . .
P21	. . .	T51621	0.18-0.22	0.20-0.40	0.20-0.40	0.20-0.30	3.90-4.25	. . .	. . .	0.15-0.25	1.05-1.25Al
Water-hardening tool steels											
W1	W108, W109, W110, W112	T72301	0.70-1.50[(e)]	0.10-0.40	0.10-0.40	0.15 max	0.20 max	0.10 max	0.15 max	0.10 max	. . .
W2	W209, W201	T72302	0.85-1.50[(e)]	0.10-0.40	0.10-0.40	0.15 max	0.20 max	0.10 max	0.15 max	0.15-0.35	. . .
W5	. . .	T72305	1.05-1.15	0.10-0.40	0.10-0.40	0.40-0.60	0.20 max	0.10 max	0.15 max	0.10 max	. . .

(a) All steels except group W contain 0.25 max Cu, 0.03 max P, and 0.03 max S; group W steels contain 0.20 max P, and 0.025 max S. Where specified, sulfur may be increased to 0.06 to 0.15% to improve machinability of group H, M, and T steels.

(b) Available in several carbon ranges.

(c) Contains free graphite in the microstructure.

(d) Optional.

(e) Specified carbon ranges are designated by suffix numbers.

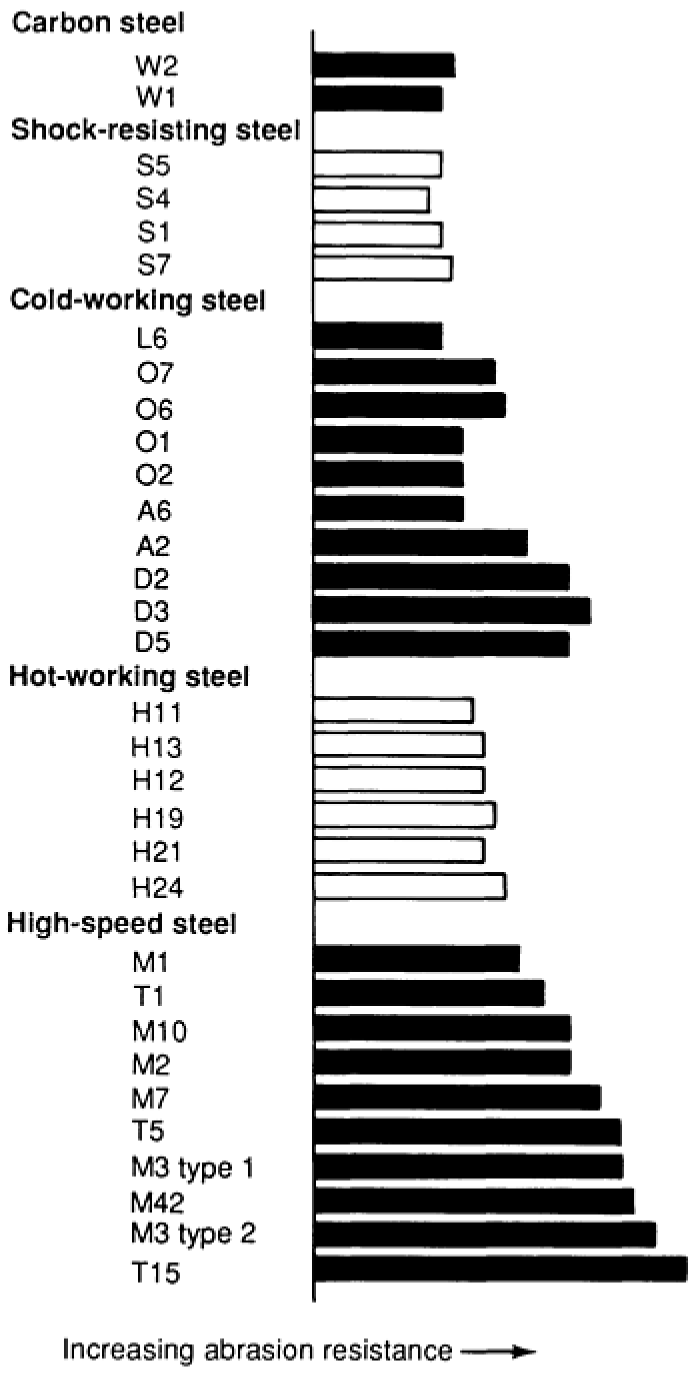


Fig. 1 Relative abrasion resistance of tool steels. Source: Ref 2

Water-hardening tool steels (Type W) contain carbon as the primary alloying element. Some grades of this type contain small amounts of either vanadium, for grain refinement, or chromium, for increased wear resistance. Although the relative wear resistance of these tool steels is low (Fig. 1), they are used in short-run wear applications where toughness is also required. Typical applications include files, woodworking tools, taps, reamers, drills, cold heading, and forming and blanking dies. These tools are heat treated to high surface hardness (HRC 59-64) before use. One of the highest-volume applications of water-hardening die steels is the coining of dies used to manufacture silverware and stainless dinnerware. The grades within this category are shallow hardening, so that when heat treated in section sizes thicker than 13 mm (0.5 in.), they have a hard case with a tough, lower-hardness core.

The abrasive wear resistance of these grades is directly related to their carbon content. As carbon content increases, so does hardness and wear resistance. This relationship is shown in Fig. 2, where linear wear resistance relative to a standard

material is plotted versus sample hardness for several levels of carbon in steel. This figure also demonstrates that abrasive wear resistance and material hardness are directly related.

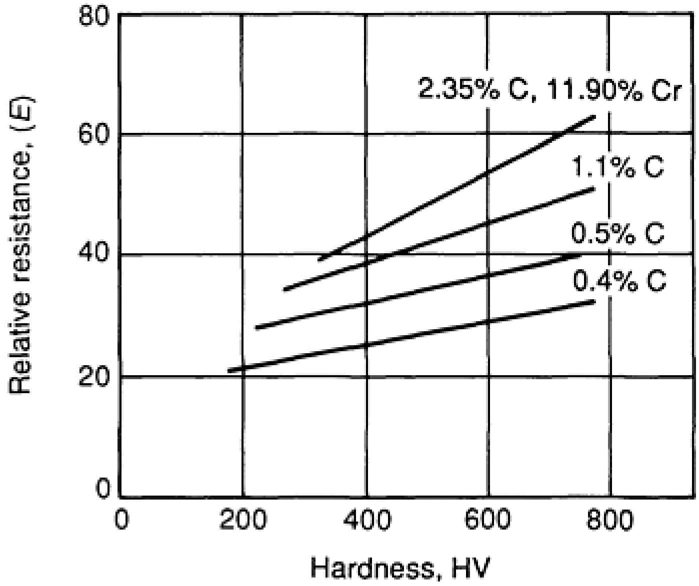


Fig. 2 Resistance to abrasive wear as function of both hardness and carbon content. Source: Ref 3

Several factors are important to the abrasion wear resistance of water-hardening tool steels: increased carbon content, increased surface hardness, and increased volume fraction of cementite. Of course, the first factor is related to both of the other factors.

Shock-resisting tool steels (Type S) have lower carbon contents than the water-hardening tool steels and are thus unable to achieve the same high hardness values. However, they are alloyed with manganese, silicon, and, sometimes, tungsten, chromium, or molybdenum to improve hardenability (depth of hardness). Vanadium sometimes is added to improve hardenability and to promote a fine grain size. Although the wear resistance of these tool materials is roughly equivalent to the water-hardening tool steels, they possess higher toughness properties. They are used in applications requiring shock resistance, such as chisels, hammers, punches, shear blades, stamps, and rivet sets.

The wear resistance of these shock-resisting steels is no longer a simple function of carbon content alone. Additions of chromium and molybdenum add not only to hardenability, but also contribute to increased wear abrasive resistance (Fig. 1). On the basis of carbon content and hardness alone, grade S5 should possess the highest wear resistance, but grades S1 and S7 both have somewhat higher wear resistance, because of their increased chromium and molybdenum contents.

Cold-work tool steels are described below in terms of three types.

Type A comprises medium-alloy, air-hardening materials intended for use at or near room temperature. Manganese, chromium, molybdenum, and vanadium are the major alloying elements that are used to promote deep hardening in these steels. Their high carbon content imparts good wear resistance in most applications, which typically include: blanking, coining, and other cold-forming tools, as well as molds, rolls, shear and slitter knives, and mandrels.

Type O comprises low-alloy steels that are hardened by oil quenching. The principal alloying elements are manganese, tungsten, molybdenum, and chromium. Applications include dies for blanking, coining, drawing, or forming, as well as gages, cams, plastic molds, shear blades, and machine ways. The oil-hardening cold-work steels are intended for room-temperature applications and rely on high carbon contents and undissolved carbides for their wear resistance.

Type D steels have high carbon and chromium contents and are deep hardening. All of the commonly used grades, except D3, contain 1 wt% molybdenum, and some contain either vanadium or cobalt. The wear resistance of these steels is exceptional and increases with increasing carbon and vanadium content. Popular applications include blanking, cold forming, lamination, and drawing dies, gages, punches, rolls, slitter and shear knives, and thread-rolling dies.

Hot-work tool steels (Type H) are medium-carbon steels with additional alloying elements that help maintain hardness and wear resistance at higher temperatures. All grades contain chromium; grades H20 to H39 contain tungsten, and H42 contains 5 wt% molybdenum. All of the hot-work steels are deeply hardenable.

Typical applications involve elevated-temperature use, as in forging and extrusion dies, bolsters, dummy blocks, gripper and hot heading dies, mandrels, punches, and piercing tools. The tungsten (H20 to H39) and molybdenum (H40 to H59) types are used in the highest-temperature applications, where resistance to softening is of the greatest importance.

High-speed tool steels are described below in terms of two types.

Type M steels have high carbon contents, with molybdenum as the major alloying element. Other alloying elements include tungsten, chromium, vanadium, and cobalt. These steels reach exceptionally high hardness after heat treatment and can maintain that hardness at elevated temperatures.

Typical applications involve metal cutting (specifically, broaches, chasers, taps, reamers, drills, end mills, hobs, lathe tools, and saws) and metalworking (punches, thread-rolling dies, cold-heading dies, and blanking dies). Woodworking applications, as well as bearing and pump components, are also candidates for high-speed tool steels.

Type T steels also have high carbon contents, with tungsten as the primary alloying element. Chromium and vanadium are present as alloying elements in all Type T steels; most also contain cobalt. These steels are used in applications similar to those that use Type M. However, the tungsten types tend to be less tough, more resistant to heat, and slightly higher in abrasion resistance.

Metallurgical Aspects of Tool Steel Wear

There are four basic mechanisms of tool steel wear: abrasion, adhesion, corrosion, and contact fatigue wear. Tool steels are most often used in metalworking processes, in which the effects of the different wear mechanisms are often combined.

Adhesive wear is usually the life-limiting wear mechanism in dry metal-forming operations. Because lubrication is very effective in alleviating adhesive wear, abrasive wear often becomes the life-limiting factor. However, wear is usually a complex mixture of the differing mechanisms. Therefore, production experience in tool steel, lubrication, and coating combinations, as well as experimentation, are the means for choosing the ideal tool steel for the application.

Abrasive wear is experienced whenever a hard phase is rubbed against tool steel. The hard phase can be embedded in a softer phase, such as oxide inclusions in steel or aluminum, or it can be in a liquid carrier, such as the grinding dust carried onto tool steel guides by the lubrication flow or the contaminants in fuel or oil that is to be pumped.

As previously discussed and shown in Fig. 2, the abrasion resistance of carbon tool steels (W type) is directly related to the hardness and carbon content of the steel. Other work by Larsen-Basse (Ref 4) shows that wear resistance also has a linear relationship with the volume fraction of cementite (Fe_3C, iron carbide), as illustrated in Fig. 3.

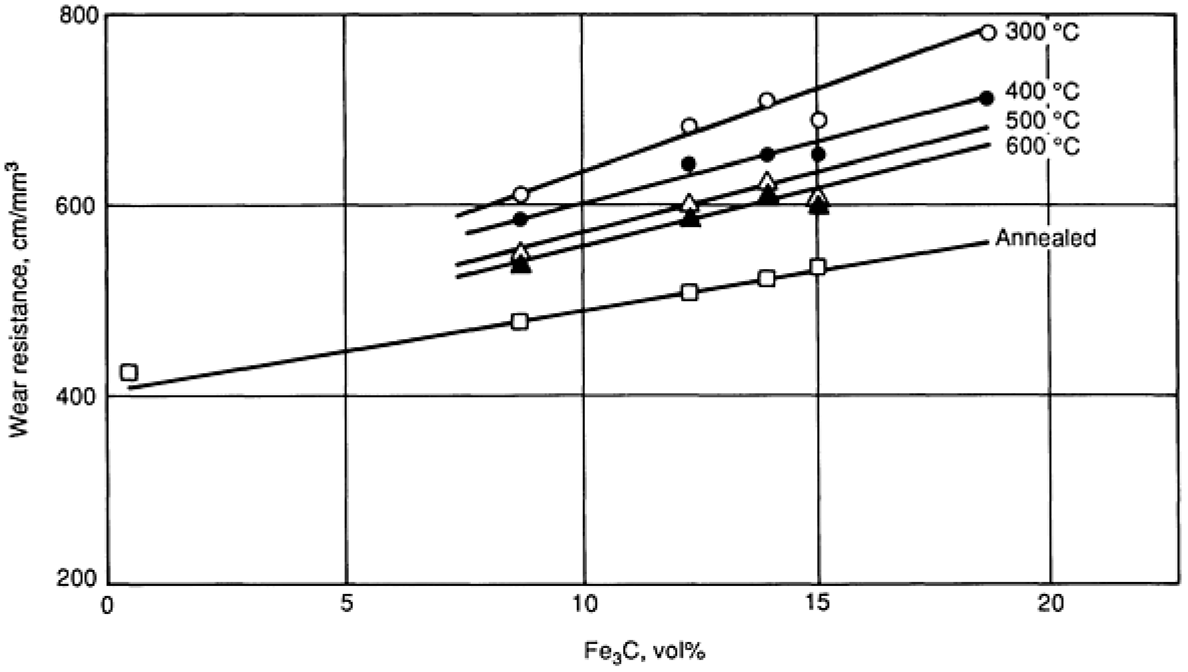


Fig. 3 Abrasion resistance of hardened and tempered carbon steels as function of volume fraction of cementite. Temperatures shown indicate tempering temperature; hardness decreases as tempering temperature increases. Source: Ref 4

Abrasion resistance was defined in Ref 4 as the sample sliding distance divided by metal removal by dry 400-grit silicon carbide sandpaper. The different tempering temperatures plotted also show the influence of hardness on wear resistance; hardness increases with decreased tempering temperature.

A detailed study of the dry abrasive wear of tool steels was conducted by Budinski (Ref 5). His assessment of the data showed very good correlation with a structural parameter, *A*, which equals the primary carbide size, multiplied by the volume fraction of carbides, multiplied by carbide hardness. The data are shown in Fig. 4. The implication of this finding is that the most abrasion-resistant tool steel will have a large volume fraction of large, hard carbides.

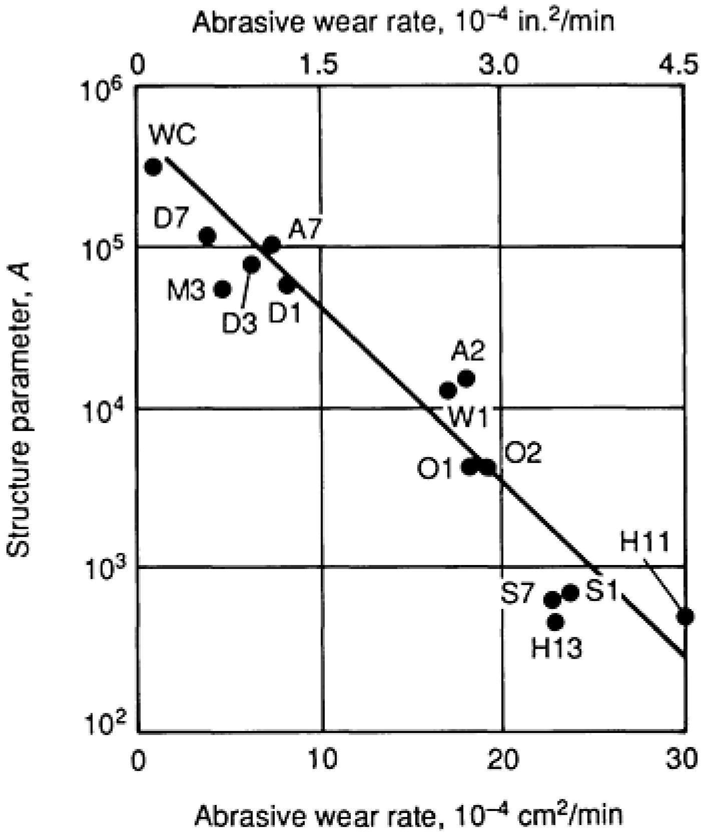


Fig. 4 Abrasion resistance of tool steels versus a structural parameter (defined in text). Source: Ref 5

Care must be used in committing these results to steelmaking practice. Most wear is metal-to-metal; therefore, abrasion is not the only mechanism operative. In addition, large carbides can be torn from the tool steel matrix and will enhance tool steel wear by adding abrasive carbide particles to the wear debris (Ref 6). This effect can also occur if tool steels are overtempered (Ref 7).

Adhesive wear between tool steel and workpiece can result when direct contact occurs because of lubricant breakdown, high pressures, or process design. This type of wear is due to the cold welding of microscopic asperities between the tool and the workpiece. The environment also has an effect. Generally, adhesive wear becomes more severe as oxygen content decreases (for example, sliding in a vacuum), because no oxide layer can form.

Metallurgical factors that reportedly decrease the adhesive wear of tool steels include higher tool steel hardness at temperature (hot hardness), the built-in lubricity of graphitic tool steels (Types O6 and A10), small grain size (Ref 8), a large volume fraction of primary carbides (Ref 9), and small primary carbide size (Ref 8). Most of these factors have been established from empirical data. However, it is not at all clear that they control adhesive wear in all cases.

For example, Budinski has evaluated the wear between unlubricated tool steel and hardened 440C steel, as well as between tool steel and annealed 1020 steel. He found no correlation between tool steel wear and certain factors, specifically carbide size, carbon content, carbide volume fraction, total alloy content, and hardness, nor did he find a structural parameter combining primary carbide size, carbide volume fraction, and hardness (Ref 5).

An extreme case of adhesive wear, sometimes called continuous wear, occurs in high-speed tool steels used in demanding metal-cutting applications. Tool temperatures at the tip and in the wear crater can approach 700 °C (1300 °F), and direct contact with the chip causes alloying elements (usually carbon and chromium) to diffuse out of the tool steel and into the chip. Adhesive wear of this severity is often called seizure, or galling. The solution is to reduce the cutting speed, coat the tool, or improve the lubrication and cooling.

Corrosive wear is usually not a major cause of excessive tool steel wear, but it can become the most important wear mechanism at high temperatures or during plastics processing. Wear proceeds by the corrosion of fresh, exposed tool steel, followed by rupture of the corrosion film by the workpiece and corrosion of the newly exposed surface. This form of wear can usually be controlled by proper lubrication, higher-chromium tool steels, or coatings.

Contact fatigue wear is caused by repeated sliding or rolling of one component over another, which leads to spalling or microcracking. This type of wear often manifests itself as cutting or punch edge chipping. Surface fatigue wear by rolling members is the subject of several other articles in this Volume and is therefore not addressed here.

Improved toughness and higher hardness are the most common metallurgical solutions to contact fatigue wear. These two properties have a strong inverse influence on each other, and care must be taken to avoid creating a larger problem when increasing either. One approach that improves toughness without influencing hardness is to decrease the carbide size by substituting a powder metallurgy product (Ref 10). Another approach is to lower the hardening (austenitizing) temperature.

Lubrication of Tool Steels

Several considerations are involved in the selection of lubrication systems for tool steel applications. The lubricant should ideally provide a continuous layer between the tool steel and workpiece in order to:

- Reduce friction and adhesive wear
- Bond to the tool steel to prevent gouging and die pickup on the workpiece
- Flush metalworking debris away from the tool steel-workpiece interface
- Extract heat away from the tool steel

Because lubrication systems are highly dependent on the operation being conducted, this discussion focuses on metalworking examples, rather than tool steel types. The coefficient of friction in metalworking processes is influenced by many variables. The values presented here only give an approximate magnitude and are just intended to illustrate the effect of lubrication on friction. Therefore, they should be used with caution.

Sheet Metal Forming. The primary function of lubrication in press forming is to prevent galling (tool material pickup). A secondary function is to reduce die wear through abrasive and chemical wear mechanisms. Light press forming of most metals requires only light mineral oil or soap lubrication to prevent die wear. More severe conditions, experienced when forming copper-, iron-, and nickel-base alloys, usually require liquid lubrication using either mineral oils with extreme-pressure (EP) additives (such as sulfur and chlorine) or zinc stearate soaps in order to prevent galling. The coefficient of friction experienced in deep drawing and stretch forming is reduced as the viscosity of the lubricant increases (Ref 11, 12, 13), with an attendant reduction in the surface quality of the formed part (Ref 11).

Solid-film lubricants in the form of polyethylene (PE), polyvinylchloride (PVC), and polytetrafluoroethylene (PTFE), which are applied to the workpiece as a powder, coating, or solid sheet, also give adequate protection against adhesive wear. Extreme conditions encountered when press forming iron- and nickel-base alloys require a phosphate-coated product and soap lubrication to prevent excessive die wear. Tin-coated steel, with additional liquid lubrication of mineral oil or emulsions to cool the die, has also been successful in preventing die pickup in severe press-forming applications.

Moderate-to-severe forming of aluminum and magnesium alloys is normally accomplished with solid film soap or wax lubrication. The forming of titanium alloys always requires at least an EP oil; as the forming becomes more severe, wax or soap lubrication is preferred. Warm forming of magnesium and titanium alloys can best be accomplished with either organic- or water-based colloidal graphite coatings of the workpiece.

Hot extrusion, particularly at higher temperatures, is an extremely demanding application for tool steels. Both adhesive and abrasive wear, in addition to thermal fatigue and high temperatures, make the choice of lubrication system critical to success. A good lubrication system should act as thermal insulation between the extrusion billet and the tool steel, promote easy sliding by providing a continuous film between the billet and tool steel, and should be chemically inert to both the extrusion tooling and billet. The coefficient of friction in hot extrusion varies between 0.015 and 0.2, which underscores the diversity of this process (Ref 14).

Lubrication systems for hot extrusion are extremely varied, because extrusion must be either totally unlubricated or fully lubricated to achieve good extrusion performance and the reduction of surface defects. The range of temperatures employed in extrusion also creates a challenge in lubrication technology. Extrusion that uses bridge, spider, or porthole dies must be carefully accomplished to ensure that no lubricant is entrapped in the pressure-welded interface created during extrusion. For this reason, and to prevent surface and subsurface entrapment of lubricant, most aluminum extrusion is conducted without lubrication.

The major problems in unlubricated aluminum extrusion are die pickup and wear. Internal water cooling of the die and carefully controlling extrusion rates to avoid excessive die temperatures are frequently used to prevent sticking and die wear. Die pickup, in this case, is primarily caused by oxygen in the form of either aluminum oxide or iron oxide. The wear resistance of the die can be improved by preventing oxygen contact with the die by (Ref 15):

- Extruding low-oxygen billet material
- Mechanically removing surface oxide from the billet by machining or shot blasting
- Maintaining an oxygen-free atmosphere around the die by shrouding with inert gas
- Cleaning or surface dressing the die between pushes to prevent die oxidation when the die is not in contact with the extruded metal
- Leaving the tail end of the extrusion piece in the die, thereby preventing oxidation of the die

Hot extrusion of copper alloys is simplified by the fact that the copper oxide formed during preheating is a good lubricant. Sometimes, the oxide lubrication is aided by spraying an oil-graphite mixture onto the container walls, mandrel, and die face. However, the oxide formed on brass alloys does not have good lubricity. Brass and hard copper alloys are therefore often extruded using glass lubrication.

For titanium, refractory metals, and iron- and nickel-base alloy extrusion, the lubrication is either graphite in oil or fat (for short pieces) or glass (Ref 16). The glass lubricant is usually a basalt mixed with borosilicate glass.

A mild steel can is sometimes welded around billets of highly alloyed stainless steel and nickel-base alloys to act as both lubricant and friction-reducing agent by preventing adhesion between the die and the extrusion billet. This process, known as either canning or jacketing, not only improves the lubrication at the interface between the tool steel die and extrusion billet, but also acts as a barrier to oxidation of the billet, thereby reducing abrasive wear of the die by hard oxide particles. Sometimes, glass lubrication is used to augment the lubrication provided by the can.

Other materials are also used for the can. The important parameters in choosing a can material are chemical inertness to both the extrusion billet and the die, lack of adhesion to the die, and a yield strength at temperature that is from one-third to one-half that of the extrusion billet yield strength. Examples are copper tubing and pure aluminum tubing over hard aluminum alloys.

Hot forging, particularly the closed-die practice, is also a very demanding application for tool steels. Lubricants used in these situations have several functions:

- Reduce friction to produce well-filled forgings at the lowest power requirements possible
- Insulate the tool steel dies from the hot work-piece
- Provide a barrier between the die and workpiece to prevent chemical attack
- Prevent erosion (abrasion) of the tool steel die
- Aid in releasing the completed forging from the die

Dry friction coefficients for uniaxial compression (upsetting) are about 0.2 for aluminum and steel and about 0.1 for copper and brass. Lubrication reduces these values to a range from 0.05 to 0.07 (Ref 17).

The most commonly used hot forging lubricants are graphite suspensions in water, grease, or oil, which are applied by spraying or swabbing. An even spray coating is preferred. Other common hot forging lubricants are sawdust (the first lubricant used); chloride salts, such as NaCl or $BaCl_2$; solid lubricants, such as talc, mica, vermiculite, white lead, and alumina; and metal films deposited on the workpiece.

Forging temperature has a pronounced influence on the type of lubrication used in forging. Most organic-based lubricants break down at temperatures above approximately 200 °C (390 °F), although aluminum stearate is apparently useful up to a temperature of about 400 °C (750 °F). Much hot forging work is done at temperatures that exceed 1000 °C (1830 °F). The only successful lubricants at these temperatures are graphite, glasses, and metallic oxides, such as copper oxide.

Cutting. The use of proper lubrication in cutting with tool steels is critical to good tool life and the prevention of wear. Cutting fluids perform several functions in metal cutting operations:

- Cooling of tool and workpiece
- Lubricating to reduce friction and adhesive wear
- Reducing built-up edge
- Flushing away the chips generated
- Protecting both the tool and workpiece from corrosion

The most important consideration is the cooling aspect, because the tool-workpiece interface temperature can rise to temperatures that exceed 600 °C (1110 °F), which causes tool softening and greatly accelerates tool wear. Proper cooling can reduce tool wear by approximately 33% (Ref 18).

The coefficient of friction in cutting depends on the material being cut, the lubricant, and, to a much lesser extent, the cutting tool steel. At low cutting speeds, the coefficient of friction appears to be a function of the number of carbon atoms per unit volume of the cutting fluid. An increasing carbon content in the lubricant results in a decreased friction coefficient. Using proper lubrication, friction can be reduced by 60% when cutting aluminum, 84% when cutting copper, and approximately 35% when cutting steel and stainless steel. Friction coefficient values are reportedly between 0.25 and 0.55 (Ref 19).

Several types of fluids are used in cutting operations. The most common are cutting oils, which are usually mineral oils or water-based miscible oils. Some synthetic fluids, gases, and solid lubricants are also used. The mineral-oil-based fluids have excellent lubricating and oxidation (rust) prevention properties. The water-based oils have better cooling properties than the mineral-oil-based products, in general, and thus are often used in high-speed machining operations.

Chemical fluids are good coolants, and some are also useful as lubricants. These fluids are chemical solutions of various inorganics dissolved in water. Although most liquid cutting fluids are applied by flooding both the tool and the workpiece, mist application is also used when visibility of the cutting operation is desired.

Air, argon, helium, and nitrogen are some common gases that are used as cutting fluids when either contamination of the workpiece or a clear view of the cutting process is necessary. Obviously, the cheapest and most common gaseous form of cutting fluid is air. Other gases with boiling points below room temperature, such as CO_2 and fluorinated hydrocarbons, are sometimes used when extra cooling is needed.

Solid lubricants are most commonly available in the form of greases, waxes, graphite compounds, molybdenum disulfide compounds, and soaps. They are most often applied by brushing in low-speed operations, such as hand tapping and reaming.

Surface Treatments

Most surface treatments are employed to increase surface hardness and/or wear resistance, minimize adhesion (reduce friction), or improve the corrosion resistance of the tool steel base.

Carburizing. The processes of case hardening and carburizing are of limited use in tool steel applications, because of the relatively high carbon contents of the tool steels. Carburizing can be accomplished in many ways, and essentially consists of heating the final machined tool into the austenite region in the presence of carbonaceous solids, liquids, or gases.

Low-carbon plastic mold steels (P type) are often carburized after hubbing or machining of the cavity in the mold. In this application, the tool steel is intentionally lean in carbon content to improved hubbing or machining, and must be carburized in order to have sufficient surface hardness for the end use.

Nitriding is a frequently used surface treatment that increases surface hardness, adds to the corrosion resistance of the tool, and reduces friction. Basically, the process involves heating the finished tool in the presence of a nitrogen-containing liquid or gas and allowing nitrogen to diffuse into the tool. Gas nitriding is usually accomplished at a lower temperature (about 527 °C, or 980 °F) and longer time (10 to 90 h) than liquid nitriding, which occurs at temperatures ranging from 538 to 552 °C (1000 to 1025 °F) for 2 to 4 h. A nitrided depth that ranges from 13 to 76 m (0.0005 to 0.003 in.) is desired.

Because of decreased wear and die pickup, cold-extrusion punches experience a two to three-fold improvement in life. Nitriding is often used whenever mold wash is a problem in the die casting of zinc or aluminum alloys. Galling of sheet metal working dies can be alleviated by nitriding these dies before use.

Steels that will be nitrided should contain one or more of the nitride-forming elements (chromium, vanadium, or aluminum) in order to prevent the easy spalling and chipping that results when iron nitride is formed. Commonly nitrided tool steels include H11, H12, H13, A2, O2, and the high-speed tool steels.

Ion or plasma nitriding has many of the same characteristics of liquid or gas nitriding. This process relies on a nitrogen gas being ionized by glow discharge conditions between the tool (cathode) and the furnace wall or shield (anode). The primary advantages are the reductions in time and temperature, which save money and reduce the distortion and softening of prehardened tools. Usually, treatment times vary between 0.5 and 36 h.

Oxidation is a well-established process used for high-speed steel-cutting tools. Increases in tool life of up to 100% are mostly due to a decrease in friction, because of the hard oxide coating and the ability of the porous oxide to entrap lubricant and draw it to the tool-work-piece interface (Ref 20). Steam oxidation of a finished tool is accomplished either by exposure to steam at a temperature of about 566 °C (1050 °F) or by treating in liquid sodium hydroxide and sodium nitrate salts at approximately 140 °C (285 °F) for periods of time ranging from 5 to 20 min. These treatments result in a black oxidized layer that is less than m (0.0002 in.) thick and will not peel, chip, nor crack, even when the tool is bent or cut.

Boriding. In this process, boron atoms from a solid, liquid, gas, or plasma atmosphere surrounding the finished part are diffused into the surface, creating a hard, wear-resistant iron boride layer. Metal-to-metal wear testing demonstrated a three-fold improvement in wear resistance of borided O1 and O2 tool steels and over a two-fold increase in A2 tool steel (Ref 21). Borided A2 tool steel showed twice the life of uncoated O2 tool steel in a deep-drawing operation in which low-carbon steel cups were manufactured (Ref 21). An H13 roller designed to flange milk cans was borided and produced three times as many cans before it wore out (Ref 22).

Boriding takes place at temperatures as low as 600 °C (1100 °F), but usual practice involves a period from 1 to 6 h at temperatures from 800 to 900 °C (1470 to 1650 °F) (Ref 21, 22). The resultant layer is between 13 and 130 m (0.0005 and 0.005 in.), and tends to be dull because of the microroughness of the surface. The high process temperature requires that the boron treatment act as the austenitization step, or else the process must be followed by reaustenitization. This necessarily limits the process to applications where tolerances of about 25 m (0.001 in.) can be tolerated.

Plating. Three different kinds of plating are used on tool steels. Cadmium plating is used for appearance purposes and to reduce corrosion of the tool. It also has some usefulness in preventing adhesion. Nickel plating is commonly used for appearance purposes and to prevent corrosion.

The most commonly practiced tool steel plating process is hard chromium plating. Plating thickness varies between 2 and 13 m (0.0001 and 0.0005 in.) and, because it is very hard, it prolongs life by increasing abrasive wear resistance. More important than plating hardness is its very low friction coefficient, which effectively prevents adhesive wear.

However, hard chromium plating is not without problems. Tool steels may be hydrogen embrittled when plated, and the plating has a tendency to spall and flake. These wear debris can actually accelerate abrasive wear.

Chemical vapor deposition (CVD), a process conducted in a vacuum chamber, relies on a deposition from reacted gas onto the tool steel surface. Many different materials can be used as coatings. Chromium, Al_2O_3, TiC, CrC, Fe_4N, and TiN are commonly used, and other materials are being studied. This process utilized high temperatures, usually above 800 °C (1472 °F), which means that tool steels must be tempered after the CVD coating is applied. The most popular wear-resistant coatings are TiC and TiN, which are used to coat high-speed, cold-work die and hot-work die tool steels. These coatings commonly range in thickness from 2 to 20 m (0.0001 to 0.001 in.). Using CVD coating with TiC and TiN, the

primary mechanism of wear reduction is the extremely high hardness, which leads to excellent abrasion resistance, although some decrease in friction coefficient can often be realized. The chlorine content of the coating must carefully be maintained at a level below 5% to avoid degradation of the wear resistance (Ref 23).

Tool steels that can be successfully CVD coated include the AISI A, S, D, H, M, and T steel types. The lower-alloyed S type and all of the W and O types are either very difficult or impossible to properly coat, because of their low austenitization temperatures.

Physical vapor deposition (PVD), which is also conducted in a vacuum chamber, can be accomplished in several different ways. The process relies on plasma-aided precipitation of either TiC or TiN onto tool steel at temperatures ranging from 200 to 550 °C (400 to 1025 °F) (Ref 23). This temperature range is much more suitable for the coating of high-speed tool steels than the temperatures required for CVD.

Tool steel wear is reduced in about the same proportions (2 to 6 times less wear), whether the TiC or TiN is applied by CVD or PVD. TiN coatings on H13 pins reduced the friction coefficient in pin-on-disk tests from 0.7 to less than 0.2 (Ref 24). Modified ASTM G65-10 to abrasive wear testing of D3 steel showed that wear of the TiN-coated samples was between 4 and 23% of the uncoated samples, depending on their initial surface roughness (Ref 24). This result led Sundquist *et al.* to propose that increases in tool life that are due to TiN coating can only be expected when the surface roughness is less than the coating thickness (Ref 24).

References

1. *Steel Products Manual, Tool Steels,* American Iron and Steel Society, 1981
2. W.H. Wills, *Matter. Eng.,* Aug 1982, p 44-45
3. M.N. Kruschov, *Proc. Conf. Lab. Wear,* Institution of Mechanical Engineers, 1957, p 655
4. J. Larsen-Basse, The Abrasion Resistance of Some Hardened and Tempered Carbon Steels, *Trans. AIME,* Vol 236, 1966, p 1461-1466
5. K.G. Budinski, Wear of Tool Steels, *Wear of Materials,* American Society of Mechanical Engineers, 1977, p 100-109
6. S. Hogmark, O. Vingsbo, and S. Fridström, Mechanisms of Dry Wear of Some Martensitic Steels, *Wear,* Vol 31, 1975, p 39-61
7. A.M. El-Rakayby and B. Mills, The Role of Primary Carbides in the Wear of High Speed Steels, *Wear,* Vol 112, 1986, p 327-340
8. P.L. Hurricks, Some Metallurgical Factors Controlling the Adhesive and Abrasive Wear Resistance of Steels, *Wear,* Vol 26, 1973, p 285-304
9. S. Söderberg and S. Hogmark, Wear Mechanisms and Tool Life of High Speed Steels Related to Microstructure, *Wear,* Vol 110, 1986, p 315-329
10. R.W. Bratt, G. Steven, and B.H. Seaton, "Particle Metallurgy of High Speed Steels," Technical Paper TE73-203, Society of Mining Engineers, 1973
11. B. Fogg, *Sheet Met. Ind.,* Vol 44, 1967, p 95-112
12. W.J. Wojtowicz, *Lubr. Eng.,* Vol 11, 1955, p 174-177
13. D.V. Wilson, *Sheet Met. Ind.,* Vol 43, 1966, p 929-944
14. S. Kalpakjian, Metal Deformation Processes, *Hot Extrusion Lubrication,* J.A. Shey, Ed., Marcel Dekker, 1970, p 565
15. R. Chadwick, Developments in Design and Application of Extrusion Presses for Metal Processing, *Int. Met. Rev.,* Vol 25 (No. 3), 1980, p 94-138
16. S. Sèjournet, *Friction and Lubrication in Metal Processing,* American Society of Mechanical Engineers, 1966, p 162-184
17. H. Takahashi and J.M. Alexander, *J. Inst. Met.,* Vol 90, 1961, p 72-99
18. M.R. Sewailem and I.M. Mobarak, *Wear,* Vol 67, 1981, p 261-269
19. M. Kronenberg, *Machining Science and Application,* Pergamon Press, 1966, p 71-77

20. T.D. Deming, Steam Treating Emerges as Important Cog in Metal Surface Engineering, *Ind. Heat,* Vol 42, 1990, p 28-30
21. H.C. Fiedler and R.J. Sieraski, Boriding Steels for Wear Resistance, *Met. Prog.,* Feb 1971, p 102-103
22. O.N. Guy, Boronizing--A Surface Heat Treatment for Critical Wear Surfaces, *New Developments in Tool Materials and Applications,* Illinois Institute of Technology, 1977
23. F. Wendl, Current Trends in Surface Treatment of Tools Used for Plastics Processing, *Thyssen Edelstahl Tech. Ber.,* special issue, May 1990, p 82-99
24. H.A. Sundquist, E.H. Sirvio, and M.T. Kurkinen, Wear of Metalworking Tools Ion Plated With Titanium Nitride, *Met. Tech.,* Vol 10, 1983, p 130-134

Selected References

- V.A. Kortesoja, *Properties and Selection of Tool Materials,* American Society for Metals, 1975
- K. Laue and H. Stenger, *Extrusion,* American Society for Metals, 1981
- *Machining Data Handbook,* Vol 1 and 2, Metcut Research Assoc., Inc., 1980
- J.A. Shey, *Tribology in Metalworking,* American Society for Metals, 1983

Friction and Wear of Sliding Bearing Materials

George R. Kingsbury, Glacier Vandervell Inc.

Introduction

A SLIDING BEARING (plain bearing) is a machine element designed to transmit loads or reaction forces to a shaft that rotates relative to the bearing. Journal bearings are cylindrical (full cylinders or segments of cylinders) and are used when the load or reaction force is essentially radial (that is, perpendicular to the axis of the shaft). Thrust bearings are ring-shaped (full rings or segments of rings) and are used when the load or reaction force is parallel to the direction of the shaft axis. Both radial and axial loads can be accommodated by flange bearings, which are journal bearings constructed with one or two integral thrust bearing surfaces. The sliding movement of the shaft surface or thrust-collar surface relative to the bearing surface is characteristic of all plain bearings. In many applications, plain bearings offer advantages over rolling-contact bearings--such as lower cost, smaller space requirements, the ability to operate with marginal lubricants, resistance to corrosion, and the ability to sustain high specific loads.

Plain-bearing principles have been employed in rotating mechanical devices since about 3000 B.C. However, the first use of a metal alloy because of its special properties for bearing service applications was probably Isaac Babbitt's adaptation of a pewter composition in 1839. The design and manufacture of metallic plain bearings as separate machinery components that are mechanically assembled in precisely machined housings are recent developments. Until early in the twentieth century, machinery bearings were machined from massive tin alloy or bronze castings, or were made by pouring molten babbitt metal into cast or machined holes in iron or steel machine members. The final bearing bore was then produced by cutting away excess material, leaving the hole with a relatively thick lining of bearing metal that was more or less completely bonded to the housing.

Since the 1930s, bearing developments have proliferated with respect to materials and material systems, mechanical design, and manufacturing technology. Single-metal cast-in-place bearings have been largely supplanted by replaceable-insert bearings made from multilayer laminated metals; as a result, the mechanical design of bearings has become an increasingly refined engineering specialty. Manufacturing technology has evolved into a remarkably sophisticated complex of metallurgical, chemical, and mechanical processes. The scale of manufacturing operations ranges from the production of small lots by job-shop methods to the extremely high-volume mass production of automotive bearings.

Sliding bearings are commonly designated by terms that describe their application, such as:

- Connecting-rod bearing
- Main bearing
- Camshaft bearing or bushing
- Electric motor shaft bushing
- Countershaft thrust washer

With respect to size, sliding bearings are frequently classified as either thin wall (that is, thinner than ~5 mm, or 0.2 in.) or heavy wall. In general, bearings with diameters >150 mm (>6 in.) are considered to be in the heavy-wall class.

Configurations can be further described as half round, full round, flanged, or washer. The Society of Automotive Engineers (SAE) standards classify thin-wall bearings into three groups:

- Sleeve-type half bearings
- Split-type bushings
- Thrust washers

Such bearings are almost always produced by high-speed forming and machining processes from materials in flat strip form, and the terms "strip-type bearing" and "sheet-metal bearing" are occasionally used to designate these components.

Heavy-wall bearings are manufactured in relatively small lots (that is, from one to a few thousand pieces) by more conventional machine shop manufacturing processes. Starting materials can be in the form of flat slabs or tubes; components made from these product forms are sometimes designated as "slab bearings" and "shell-cast bearings," respectively.

Sliding bearings are also frequently classified according to material construction, as single-metal (solid), bimetal (two-layer), or trimetal (three-layer) bearings. These terms refer to the number of principal functional layers that are used. Each one of these types of construction is in widespread commercial use and in a wide variety of applications. The use of two or three separate layers provides a means for developing property combinations that cannot be obtained with single-metal bearings.

From the standpoint of bearing material design and selection, some or all of the following wear damage mechanisms must be considered:

- Surface fatigue wear
- Abrasive wear
- Adhesive wear
- Erosive wear
- Corrosion

Typical examples of damage caused by these processes are shown in Fig. 1.

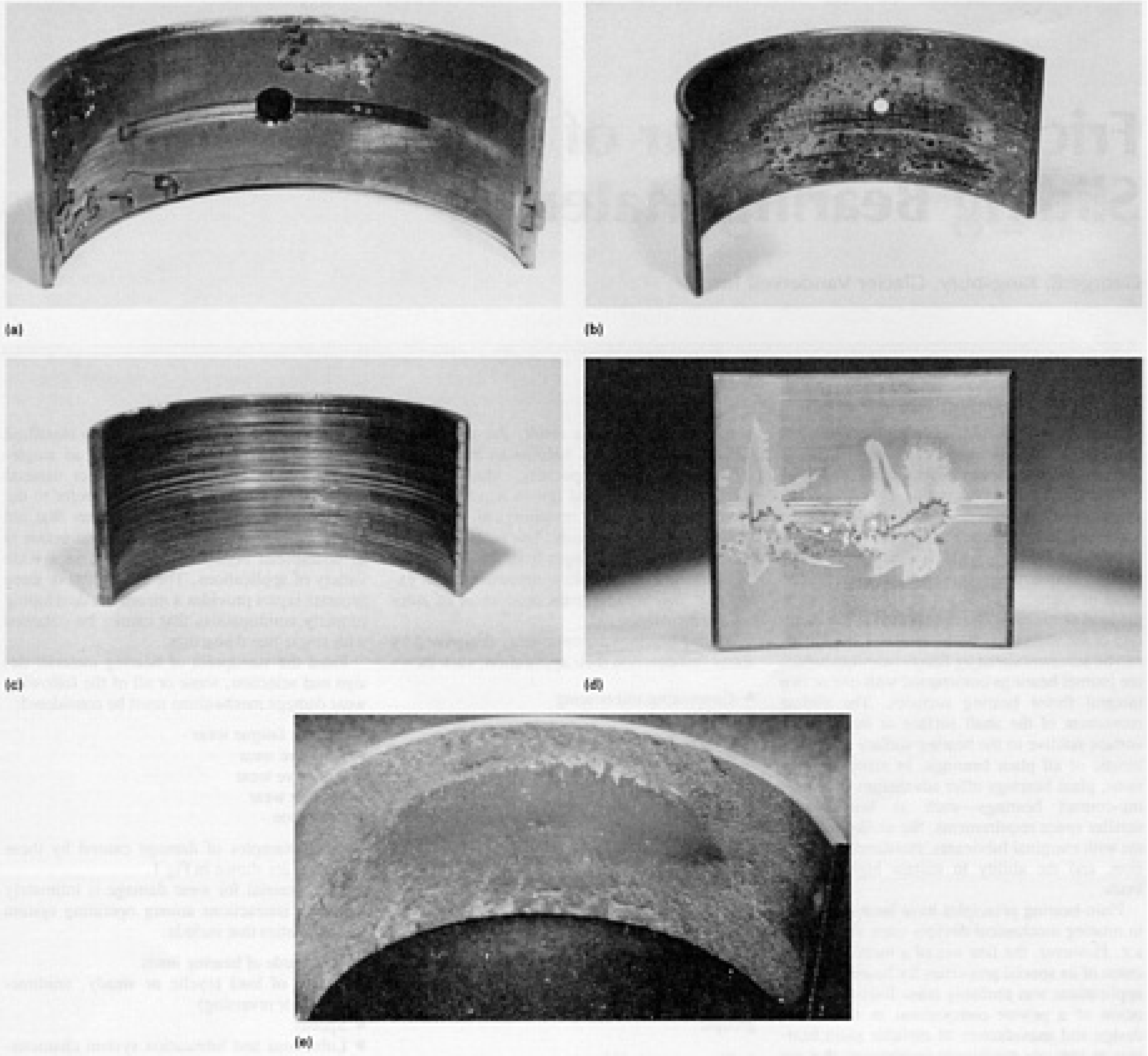


Fig. 1 Typical damage in sliding bearings caused by different wear processes. (a) Surface fatigue. (b) Abrasive wear. (c) Adhesive wear (scoring). (d) Erosive wear (cavitation). (e) Corrosion

The potential for wear damage is intimately linked to interactions among operating system characteristics that include:

- Magnitude of bearing loads
- Nature of load (cyclic or steady, unidirectional or reversing)
- Speeds
- Lubricants and lubrication system characteristics
- Lubricant cleanliness and filtration
- Operating temperatures
- Counterface (shaft) material and finish
- Alignment and rigidity
- Life expectancy

Additional information is available in the article "Friction and Wear of Sliding Bearings" in this Volume.

Properties of Bearing Materials

Surface and Bulk Properties. The conditions under which plain bearings must operate and the wide ranges over which these conditions can vary lead to concern about two kinds of bearing material properties: surface properties (those associated with the bearing surface and immediate subsurface layers) and bulk properties.

Conventional engineering definitions of materials properties do not adequately describe the distinctive properties that are needed in materials for sliding bearings. Table 1 lists some of the more important relationships between bearing material properties and wear damage mechanisms.

Table 1 Influence of bearing material properties on damage mechanisms

Bearing material property	Damage mechanisms counteracted	Causes
Compatibility	Adhesive wear	Asperity contact, surface roughness, inadequate oil film
Conformability	Adhesive wear, surface fatigue wear	Misalignment, edge loading, faulty assembly, geometric errors
Embeddability	Abrasive wear, adhesive wear	Dirt particles in oil, scoring
Fatigue strength	Surface fatigue wear	Excessive dynamic load, localized overload (for example, misalignment)
Hardness	Extrusion, erosive wear	Overheating, excessive load, cavitation effects
Corrosion resistance	Corrosive wear	Lubricant deterioration (overheating), lubricant contamination (water, fuel, refrigerant, combustion products)

Compatibility can be regarded as a purely surface characteristic. Conformability and embeddability involve the surface and immediate subsurface, and are strongly related to the bulk properties of strength and hardness. The other characteristics relate principally to bulk properties.

Measurement and Testing. Of the six characteristics listed in Table 1, only hardness can be measured satisfactorily by standard laboratory test methods. Many special dynamic test rigs and test methods have been developed by the plain bearing industry to evaluate and to measure the other characteristics and their interactions. Most of these testing systems are designed to subject specimen bearings to operating conditions that are qualitatively similar to the conditions of intended service, but quantitatively more severe with respect to one or several key parameters, such as:

- Magnitude of load
- Surface speed
- Operating temperature
- Lubricant supply
- Lubricant cleanliness

Figure 2 shows two different kinds of machines that are widely used to test engine crankshaft bearing specimens. Accelerated fatigue strength tests conducted in these machines can produce failures in $<10\%$ of the normal bearing life in a test engine.

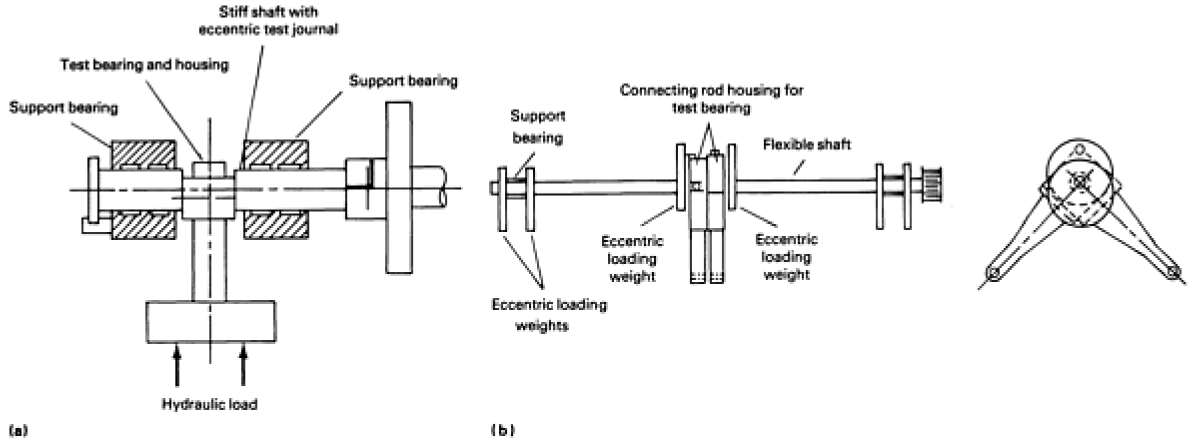


Fig. 2 Schematics showing key components of two types of engine bearing test machines. (a) "Sapphire" hydraulically loaded machine. (b) Front (left) and side (right) views of Underwood centrifugally loaded machine

Although much useful information has been developed through laboratory rig testing, it still is often necessary to test bearing materials and designs in full-size operating machines in order to clearly establish their overall suitability. Such testing is necessarily expensive and time-consuming. It should be undertaken only after careful study both of the conditions under which the bearing will operate and of prior experience in similar applications.

An appreciation of the relative importance of the various bearing material characteristics in the application at hand and an understanding of the trade-offs that are available usually will limit the viable choices to a manageable number. Full-scale bearing tests will then serve more for confirmation of the validity of design and material selections than for the exploration of design and material variables.

Compatibility. Compatibility is an antiseizure/antiscore characteristic that relates principally to the ease with which the bearing material surface will adhere or weld to a steel or iron journal surface under the influence of pressure and heat, and in the absence of a lubricant or some other intervening surface film. Some potential for scoring and seizure exists under all boundary and thin-film lubrication conditions.

Elemental metals that exhibit good compatibility with ferrous journal surfaces are all b-subgroup metals, with effective atomic radii substantially greater than that of iron. Only seven of these metals are commercially significant: silver, cadmium, indium, tin, antimony, lead, and bismuth. Of these, tin and lead offer the most attractive combinations of cost, availability, and engineering properties. Both metals are widely employed in bearing alloys because of their contributions to compatibility--either as the alloy base as in lead and tin babbitts, or as major alloy constituents as in leaded bronzes and aluminum-tin alloys.

Useful compatibility with steel is also exhibited by many nonmetals, including synthetic resins, carbon (if iron is in α phase), cemented carbides, intermetallic compounds, and ceramics. For all of these materials, compatibility probably can be attributed to a substantial lack of unsatisfied surface-atom force fields.

Conformability and Embeddability versus Hardness and Fatigue Strength. Both conformability and embeddability depend on the yielding and the plastic flow of the bearing material under load. Thus, it is not surprising that soft, weak, low-modulus metals such as tin and lead exhibit the best conformability and embeddability. Conversely, high fatigue strength is exhibited by harder, stronger, higher-modulus metals such as copper and aluminum. Useful compromises between these sets of opposed properties are effected by alloying to produce polyphase structures with intermediate properties, and by employing layered constructions in which softer and weaker surface layers are reinforced by one or more harder, stronger backing layers.

Bearing Material Microstructures. All commercially significant bearing metals, with the exception of silver, are polyphase alloys. As indicated below, these polyphase alloys can be classified according to basic microstructural types.

Type 1: Soft Matrix with Discrete Hard Particles (Fig. 3). Lead and tin babbitts are of this type. The compatibility, conformability, and embeddability of these alloys are somewhat lower than those of unalloyed lead or tin due to the presence of hard intermetallic and metalloid particles, which effectively increase bulk strength properties.

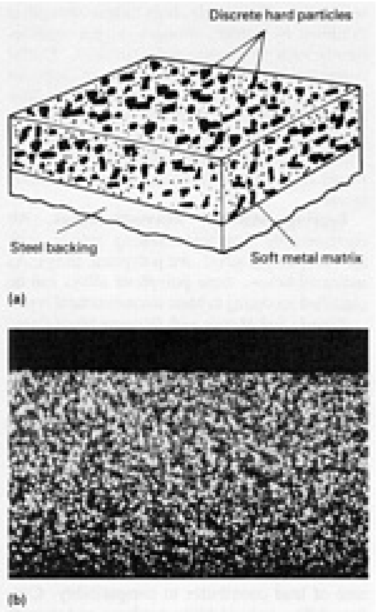


Fig. 3 Type I microstructure present in lead and tin babbitt bearing alloys. (a) Schematic showing discrete hard particles dispersed in a soft matrix that is bonded to a steel backing. (b) Photomicrograph of a lead-base babbitt in which lead is present in the dark phase and antimony in the light phase. 50×

Type II: Interlocked Soft and Hard Continuous Phases (Fig. 4). Many copper-lead and leaded bronze alloys are of this type. These structures consist of continuous mutually supporting copper and lead sponges. A large volume of lead contributes to compatibility. Conformability, embeddability, hardness, and strength properties are intermediate between those of lead and those of copper.

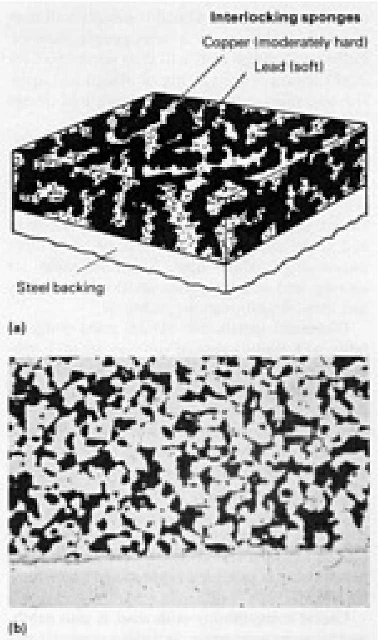


Fig. 4 Type II microstructure present in numerous copper-lead and leaded bronze bearing alloys. (a) Schematic showing interlocking soft-phase (lead) constituent and hard-phase (copper) constituent in a spongelike

structure that is bonded to a steel backing. (b) Photomicrograph of a copper-lead babbitt in which lead is present in the dark phase and copper is present in the light phase. 50×

Type III: Strong Matrix with Discrete Soft-Phase Pockets (Fig. 5). Low-lead bronzes and some aluminum-tin alloys are of this type. These structures consist of a continuous copper-base or aluminum-base metallic matrix that contains discrete pools or pockets of lead or tin. Conformability, embeddability, strength, and hardness properties are dictated by the strength of the matrix phase. Compatibility is enhanced by the soft metal phase that is exposed at the bearing surface.

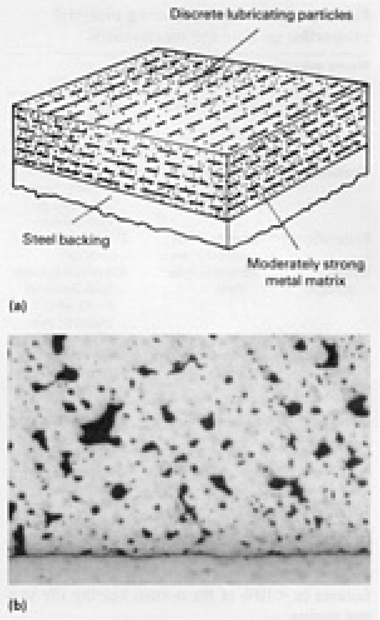


Fig. 5 Type III microstructure commonly found in low-lead bronzes and some aluminum-tin bearing alloys. (a) Schematic showing discrete soft particles dispersed in a hard matrix that is bonded to a steel backing. (b) Photomicrograph of a medium-lead tin bronze in which lead is present in the dark phase and copper-tin is present in the light phase. 50×

Mixed microstructures that combine the characteristics of types I and III are exhibited by aluminum-silicon-tin alloys and aluminum-silicon-lead alloys. In these alloys, the aluminum matrix contains both dispersed hard particles (silicon) and soft-phase (lead or tin) particles.

One of the most useful concepts in bearing material design came in 1941 with the recognition that the effective load capacities and fatigue strengths of lead and tin alloys are sharply increased when these alloys are used as thin layers intimately bonded to strong bearing backs of bronze or steel. Use is made of this principle (Fig. 6) in two-layer constructions (Fig. 7) using a surface layer of lead or tin alloy, usually no more than 0.13 mm (0.005 in.) thick. Unimpaired compatibility is provided by such a layer, together with reasonably high levels of conformability and embeddability. Other useful compromises can be effected between surface and bulk properties by employing an intermediate copper or aluminum alloy layer between the surface alloy layer and a steel back (Fig. 8). In these three-layer constructions, use of a surface layer thickness as low as 0.013 mm (0.0005 in.) offers even more favorable compromises between surface and bulk properties than are possible with two-layer constructions.

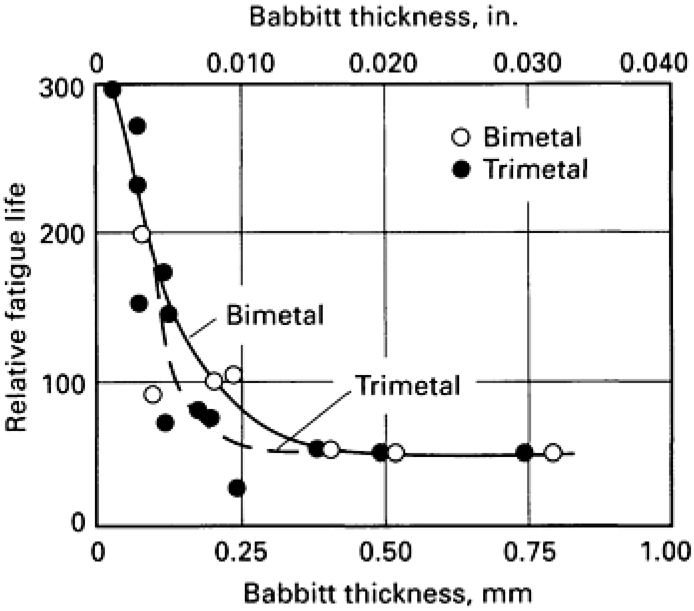


Fig. 6 Plot of bearing fatigue life versus thickness of a lead alloy babbitt surface layer as a function of the number of layers used. Bearing load, 14 MPa (2 ksi)

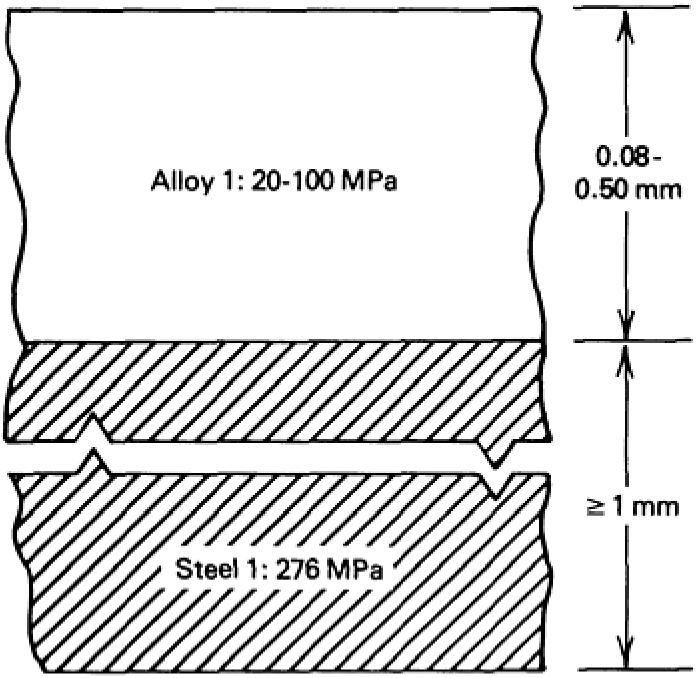


Fig. 7 Compressive yield strength levels typically obtained in the individual layers of a bilayer construction bearing material

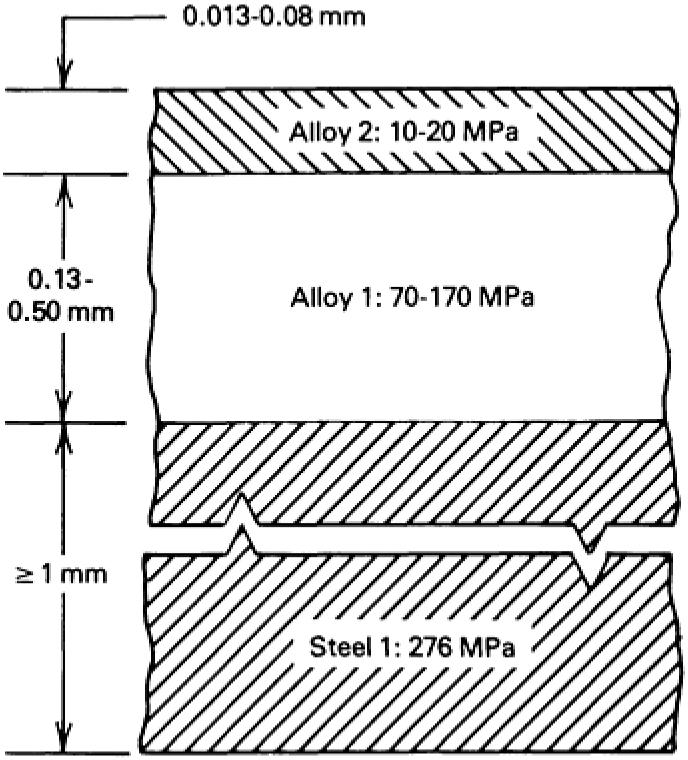


Fig. 8 Compressive yield strength levels typically obtained in the individual layers of a trilayer construction bearing material

Corrosion Resistance. Bearing failure due to corrosion alone is rare. Corrosion usually interacts with mechanical and thermal factors to produce failure by fatigue or seizure under conditions the bearing normally would be able to tolerate. To a considerable extent, bearing corrosion can be avoided by use of oxidation inhibitors in commercial lubrication oils, and by periodic oil changes. There are, however, many situations in which neither of these practices is dependable and where bearing materials with inherently high corrosion resistance should be used.

Commercially pure lead is susceptible to corrosion by fatty acids. Lead-base and copper-lead bearing alloys can suffer severe corrosion damage in acidic lubricating oils. Tin additions in excess of ~5% provide effective protection against this kind of corrosion, and for this reason tin is used extensively in lead-base bearing alloys. Both copper and lead are attacked by acidic oils that contain sulfur. This is of particular concern with copper-lead and leaded bronze bearing alloys. Effective protection can be obtained by employing layered construction, with a surface layer of either a lead alloy containing tin or a tin alloy. As long as the corrosion-resistant surface layer is intact, the underlying copper-lead alloy will not suffer damage by corrosion.

Tin and aluminum bearing alloys are substantially impervious to corrosion by the products of oil oxidation, and they are used extensively in applications where the potential for lubricating oil corrosion is known to be high. Although lubricating oil oxidation and contamination are the most common causes of bearing damage by corrosion, other sources of bearing corrosion, such as seawater, animal and vegetable oils, and corrosive gas, should be recognized. Selection and specification of a bearing material for a specific application should take into account the anticipated service conditions under which the bearings will have to operate, and the potential for corrosion that these conditions may stimulate.

Heat and Temperature Effects. The reduced mechanical strength of bearing liner materials at elevated temperatures is an important consideration in the selection of a bearing material for a given application. Fatigue strength, compressive yield strength, and hardness decrease significantly with increased bearing operating temperature. As shown by the softening curves in Fig. 9, lead-base and tin-base bearing alloys are most severely limited in this respect, and copper alloys the least.

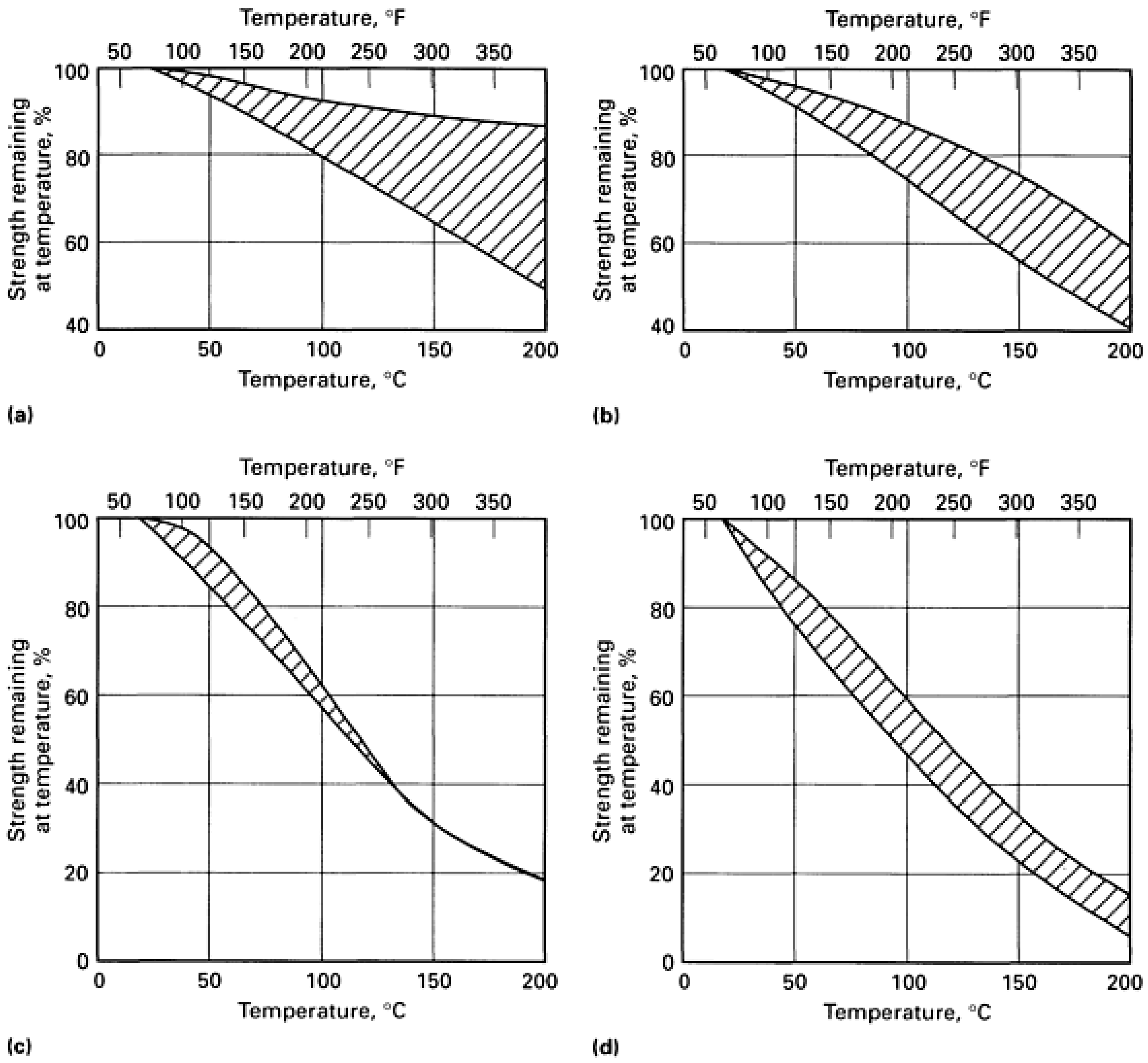


Fig. 9 Strength retention at elevated temperatures for selected bearing alloys. (a) Copper-base alloys. (b) Aluminum-base alloys. (c) Zinc-base alloys. (d) Lead-base alloys and tin-base alloys

Load Capacity. The load capacity of a bearing material is defined as the maximum unit pressure under which the material can operate without excessive friction or wear damage. Load capacity ratings published as guides for machinery designers generally represent upper limits, which may be safely employed only under very good conditions of lubricant film integrity, counterface finish, mechanical alignment, and temperature control.

In cyclic-load service (for example, in crankshaft bearings), load capacity is primarily limited by fatigue strength. In steady-load service, it depends more strongly on compressive yield strength, reflected in indentation hardness. In all cases, the strength of the material at operating temperature will be the determining factor that governs the choice of bearing material. Temperature and its control are therefore critically important to the successful operation of sliding bearings.

Although useful to the designer as reference values, load capacity ratings must be recognized as imprecise and somewhat judgmental approximations. They are not guaranteed or directly measurable material properties.

Bearing Material Systems

Because of the widely varying conditions under which bearings must operate, commercial bearing materials have evolved as specialized engineering materials systems rather than as commodity products. They are used in relatively small tonnages and are produced by a relatively small number of manufacturers. Much proprietary technology is involved in

alloy formulation and processing methods. Successful selection of a bearing material for a specific application often requires close technical cooperation between the user and the bearing producer.

Single-Metal Systems

Most single-metal sliding bearing are made from either copper alloys or aluminum alloys. Some use is also made of cast zinc-base alloys which serve as lower-cost substitutes for solid bronze. Commercially significant alloys that are used as single-metal bearings are listed in Table 2.

Table 2 Single-metal bearing material systems

Class	Material	Bearing performance characteristics[(a)]					Load capacity rating[(c)]		Typical applications
		Compatibility	**Conformability**	**Embeddability**	**Fatigue strength**	**Corrosion resistance[(b)]**	**MPa**	**ksi**	
1	Commercial bronze (10% Zn)	F	E	F	D	B	28	4	Bushings, washers
2	Tin bronze High lead (16-25% Pb)	D	D	D	D	E	21	3	Mill-machinery bearings, pump bearings, railroad-car bearings
	Medium lead (4-10% Pb)	E	E	E	C	D	28	4	Wrist pin bushings, pump bushings, electric-motor bushings, track-roller bushings, farm-equipment gear bushings, mill-machinery bearings, machine-tool bearings
	Low lead (1-4% Pb)	F	F	F	B	B	34	5	Wrist pin bushings, mill-machinery bearings, machine-tool bearings, earth-moving machinery bearings, farm-equipment gear bushings
	Unleaded	F	F	F	A	B	34	5	Wrist pin bushings, mill-machinery bearings, machine-tool bearings, railroad-car wheel bearings
3	Aluminum alloy, low tin	D	D	D	D	A	28	4	Connecting-rod main bearings, bushings, mill-machinery bearings
4	Zinc alloy 12% Al	E	E	F	B	E	28	4	Compressor bearings, pump bushings, mill-machinery bearings, earth-moving machinery bearings
	27% Al	E	F	F	A	E	34	5	Compressor bearings, pump bushings, mill-machinery bearings, earth-moving machinery bearings
5	Porous metal Bronze	C	C	C	D	B	14	2	Electric motor bushings, home appliance bearings, agricultural equipment bushings
	Iron	D	D	C	D	B	21	3	Electric motor bushings, home appliance bearings, agricultural equipment bushings
	Iron-bronze	D	D	C	D	B	21	3	Electric motor bushings, home appliance bearings, agricultural equipment bushings

(a) Bearing performance characteristics rated on scale A through F, where A is highest (best) and F is lowest (poorest).

(b) Corrosion resistance refers to corrosion by fatty acids of the kind that can form in petroleum-base oils.

(c) Load capacity rating approximates maximum safe unit loading for operation with steel journal under cyclic loading and excellent lubrication.

Wide ranges of compositions and properties are available in the older copper group. Brasses and bronzes have been widely used in bearing applications since the mid-1800s. Interest in the use of aluminum alloys was stimulated by World War II metal shortages and greatly accelerated by the commercial introduction of aluminum-tin bearing alloys in 1946. Since then, metal economics have dictated the use of aluminum alloy bearings, but brasses and bronzes continue to be preferred by many designers of heavy and special-purpose machinery.

Single-metal systems do not exhibit outstandingly good surface properties, and their tolerance of boundary and thin-film lubrication conditions is limited. As a result, the load capacity rating for a single-metal bearing usually is low relative to the fatigue strength of the material from which it was made. Because of their metallurgical simplicity, these materials are well suited for small-lot manufacturing from cast tubes or bars, using conventional machine shop processes.

Copper Alloys. Except for commercial bronze and low-lead tin bronze, copper alloys in single-metal systems are almost always used in cast form. This provides thick bearing walls (≥3.20 mm, or 0.125 in.) that are strong enough so that the bearing is retained in place when press fitted into the housing.

Commercial bronze and medium-lead tin bronze alloys C83420 and C83520 are used extensively in the form of wrought strip for thin-wall bushings, which are made in large volumes by high-speed press forming. The relatively poor compatibility of these alloys can be improved by embedding a graphite-resin paste in rolled or pressed-in indentations, so that the running surface of the bushing consists of interspersed areas of graphite and bronze. Such bushings are widely used in automotive engine starting motors.

The lead in leaded tin bronzes is present in the form of free lead that is dispersed throughout a copper-tin matrix so that the bearing surface consists of interspersed areas of lead and bronze. In general, the best selection of materials from this group for a given application will be the highest-lead composition that can be used without risking excessive wear, plastic deformation, or fatigue damage.

Aluminum Alloys. Virtually all solid aluminum bearings used in the United States are made from alloys containing from 5.5 to 7% Sn, plus smaller amounts of copper, nickel, silicon, and magnesium. Starting forms for bearing fabrication include cast tubes as well as rolled plate and strip, which can be press formed into half-round shapes. As is the case with solid bronze bearings, relatively thick bearing walls are employed in solid aluminum alloy bearings.

The tin in these alloys is present in the form of free tin that is dispersed throughout an aluminum matrix so that the bearing surfaces consist of interspersed areas of aluminum and tin. Surface properties are enhanced by the free tin in much the same way that those of bronze are improved by the presence of free lead.

The high thermal expansion of aluminum poses special problems in maintaining press fit and running clearances. Various methods are employed for increasing yield strength (for example, heat treatment and cold work) to overcome plastic flow and permanent deformation under service temperatures and loads.

Zinc Alloys. During the past 20 years, zinc-aluminum-copper casting alloys have been used to replace cast bronze alloys in certain low-speed machinery bearing applications. This practice has advanced farthest in Europe, as an outgrowth of World War II material substitution efforts.

These alloys do not contain any soft microconstituents that correspond to the lead used in bearing bronzes and to the tin in cast aluminum bearing alloys. To a considerable degree, compatibility of the zinc-base alloys seems to derive from their chemical behavior with hydrocarbon lubricants. Formation of a stable low shear strength film of zinc-base soap appears to be an important factor.

Porous Metal Bushings. Oil-impregnated porous metal bushings can also be included in the single-metal systems category. The materials used for these bushings include unleaded and leaded tin-bronze, bronze-graphite, iron-carbon, iron-copper, and iron-bronze-graphite compositions. Oil content of these materials constitutes 8 to 30% of total volume.

Bimetal Systems

All bimetal systems employ a strong bearing back to which a softer, weaker, relatively thin layer of a bearing alloy is metallurgically bonded. Low-carbon steel is by far the most widely used bearing-back material, although alloy steels, bronzes, brasses, and (to a limited extent) aluminum alloys are also used. The bimetal bearing material systems currently in significant commercial use are classified in Table 3.

Table 3 Bimetal bearing material systems

Class	Backing layer	Surface layer	Bearing performance characteristics[a]					Load capacity rating[c]		Typical applications
			Compatibility	**Conformability**	**Embeddability**	**Fatigue strength**	**Corrosion resistance[b]**	**MPa**	**ksi**	
1	Steel	Tin babbitt:								
		0.25-0.50 mm (0.010-0.020 in.)	A	A	A	F	A	14	2	Connecting-rod and main bearings, camshaft bearings, electric-motor bushings, pump bushings, thrust washers
		0.102 mm (0.004 in.)	A	B	B	E	A	17	2.5	
2	Steel	Lead babbitt:								
		0.25-0.50 mm (0.010-0.020 in.)	A	A	A	F	B	14	2	Connecting-rod and main bearings, camshaft bearings, transmission bushings, pump bushings, thrust washers
		0.102 mm (0.004 in.)	A	B	B	E	B	17	2.5	
3	Steel	Aluminum alloy:								
		High tin	B	C	C	D	A	41	6	Connecting-rod and main bearings, camshaft bearings, transmission bushings, pump bushings, thrust washers
		Medium-tin	B	C	C	C	A	55	8	Connecting-rod and main bearings, camshaft bearings, transmission bushings, pump bushings, thrust washers
		High-lead	B	C	C	C	A	55	8	Connecting-rod and main bearings, camshaft bearings, transmission bushings, pump bushings, trust washers
		Low-tin	D	D	D	C	A	55	8	Camshaft bearings, transmission bushings, thrust washers
		Tin-free	D	D	D	C	A	55	8	Camshaft bearings, transmission bushings, thrust washers
4	Steel	Copper alloy:								
		Copper-lead	C	C	C	C	F	38	5.5	Connecting-rod and main bearings, camshaft bearings
		High-lead bronze	D	D	D	C	E	45	6.5	Camshaft bearings, turbine bearings, pump bushings, thrust washers
		Medium-lead bronze	E	E	E	B	D	55	8	Piston pin bushings, rocker-arm bushings, wear plates, steering-knuckle bushings, guide bushings, thrust washers
5	Medium-lead bronze	Tin babbitt, 0.25-0.50 mm (0.010-0.020 in.)	A	A	A	F	B	14	2	Connecting-rod and main bearings, thrust washers, railroad-car journal bearings, mill-machinery bearings
6	Medium-lead bronze	Lead babbitt: 0.25-0.50 mm (0.010-0.020 in.)	A	A	A	F	C	14	2	Connecting-rod and main bearings

7	Medium-lead-bronze	Lead babbitt: mm (0.001 in.)	0.025	A	C	B	C	C	48	7	Connecting-rod and main bearings
8	Aluminum alloy, low tin	Lead babbitt, mm (0.001 in.)	0.025	A	C	B	D	C	41	6	Connecting-rod and main bearings

(a) Bearing performance characteristics rated on scale A through F, where A is highest (best) and F is lowest (poorest).

(b) Corrosion resistance refers to corrosion by fatty acids of the kind that can form in petroleum-base oils.

(c) Load capacity rating approximates maximum safe unit loading for operation with steel journal under cyclic loading and excellent lubrication.

The strengthening effect of a steel bearing back is illustrated clearly for classes 3 and 4 in Table 3; these ratings can be compared with those for the aluminum and copper alloy single-metal systems in Table 2. When steel bearing backs are employed, load-capacity ratings for both copper and aluminum alloys are sharply increased above those of the corresponding single metals without degrading any other properties. Similarly, in classes 1, 2, 5, 6, and 7, the strong bearing-back materials permit use of lead and tin alloys that have extremely good surface properties but that are so low in strength that they can be used as single-metal bodies only under very light loads.

The strengthening effect of thin-layer construction on lead and tin alloys is illustrated in Table 3 (classes 1 and 2), where a 50% increase in load capacity is achieved by reducing babbitt layer thickness. Although similar behavior has been observed with aluminum and copper alloys, the thin-liner effects are less dramatic. Liner thicknesses employed with these stronger alloys are established by metal economics and manufacturing process considerations, rather than by strength/thickness relationships.

Deterioration in surface properties with increasing liner alloy fatigue strength is clearly seen by the comparison of classes 1 and 2 with classes 3 and 4, and by comparisons within classes 3 and 4 (Table 3). In practice, only those systems with surface properties rated "D" or better are successful under boundary and thin-film lubrication conditions. This restricts the use of bimetal materials in connecting-rod and main bearings to loads of $\leq$55 MPa ($\leq$8 ksi).

Bronze-back bearings (see Table 3, classes 5, 6, and 7) do not exhibit combinations of performance characteristics substantially different from those of steel-back bearings. The practical advantages of bronze as a bearing-back material lie partly in the economics of small-lot manufacturing and partly in the relative ease with which worn bronze-back bearings can be salvaged by rebabbitting and remachining. From the standpoint of performance, the advantage of bronze over steel as a bearing-back material is the protection bronze affords against catastrophic bearing seizure in case of severe liner wear or fatigue. Similar protection is provided by the aluminum alloy bearing back in class 8.

Although the surface properties of bronze bearing-back materials are not impressive, they are superior to those of steel, and these "reserve" bearing properties can be of considerable practical importance in large expensive machinery used in certain critical applications.

Trimetal Systems

Virtually all trimetal systems employ a steel bearing back, an intermediate layer of relatively high strength, and a tin alloy or lead alloy surface layer. The systems in current commercial use are listed by classes in Table 4. Most of these systems are derived from the bimetal systems of Table 3 (classes 3 and 4) by the addition of a lead-base or tin-base surface layer.

Table 4 Trimetal bearing material systems

Class	Backing layer	Intermediate layer	Bearing performance characteristics[(a)]						Load capacity rating[(c)]		Typical applications
			Surface layer	**Compatibility**	**Conformability**	**Embeddability**	**Fatigue strength**	**Corrosion resistance[(b)]**	**MPa**	**ksi**	
1	Steel	Medium-lead bronze	Tin babbitt, 0.25-0.50 mm (0.010-0.020 in.)	A	A	A	F	B	14	2	Large connecting-rod and main bearings, bushings
2	Steel	High-lead bronze	Tin babbitt, 0.25-0.50 mm (0.010-0.020 in.)	A	A	A	F	B	14	2	Large connecting-rod and main bearings, bushings
3	Steel	Copper-lead	Lead babbitt, 0.075 mm (0.003 in.)	A	B	B	E	C	21	3	Connecting-rod and main bearings, camshaft bearings
4	Steel	Copper-lead	Lead babbitt, 0.025 mm (0.001 in.)	A	C	C	B	C	59	8.5	Connecting-rod and main bearings, bushings
5	Steel	High-lead bronze	Lead babbitt, 0.025 mm (0.001 in.)	A	C	C	B	C	83	12	Connecting-rod and main bearings, thrust washers
6	Steel	Medium-lead bronze	Lead babbitt, 0.025 mm (0.001 in.)	A	D	D	A	C	83	12	Connecting-rod and main bearings
7	Steel	Aluminum, low tin	Lead babbitt, 0.025 mm (0.001 in.)	A	C	C	B	B	55	8	Connecting-rod and main bearings
8	Steel	Aluminum, tin free, low alloy	Lead babbitt, 0.025 mm (0.001 in.)	A	C	C	B	B	55	8	Connecting-rod and main bearings
9	Steel	Aluminum, tin free, low alloy, precipitation hardened	Lead babbitt, 0.025 mm (0.001 in.)	A	C	C	B	B	76	11	Connecting-rod and main bearings
10	Steel	Aluminum, tin free, high alloy	Lead babbitt, 0.025 mm (0.001 in.)	A	C	C	B	B	76	11	Connecting-rod and main bearings
11	Steel	Silver	Lead babbitt, 0.025 mm (0.001 in.)	A	D	D	A	B	83	12	Connecting-rod and main bearings for aircraft reciprocating engines

(a) Bearing performance characteristics rated on scale A through F, where A is highest (best) and F is lowest (poorest).

(b) Corrosion resistance refers to corrosion by fatty acids of the kind that can form in petroleum-base oils.

(c) Load capacity rating approximates maximum safe unit loading for operation with steel journal under cyclic loading and excellent lubrication.

The strengthening effects of thin-layer construction are notable in those systems that incorporate electroplated lead alloy surface layers ≤0.025 mm (≤0.001 in.) thick (Table 4, classes 4 through 11). Comparison of fatigue strength and load capacity ratings of these systems with those of the corresponding bimetal systems in Table 3 shows that the thin lead alloy surface layer upgrades not only surface properties but also fatigue strength. The increase in fatigue strength can be attributed at least in part to the elimination of stress raisers, from which fatigue cracks can propagate.

Class 1 and class 2 trimetal systems comprise leaded bronze intermediate layers with relatively thick tin alloy surface layers. These represent an evolution from bronze-back babbitt construction wherein steel has replaced most of the bronze. This produces the expected economy and bearing-back yield strength, but retains the desirable "reserve" bearing properties exhibited by bronze-back construction.

Class 11 trimetal systems, which have silver intermediate layers, are too costly for most commercial applications. However, they provide an unequaled combination of high load capacity and corrosion resistance. They continue to have limited use in radial piston engines for aircraft.

Trimetal systems with electroplated lead-base surface layers and copper or aluminum alloy intermediate layers provide the best available combinations of cost, fatigue strength, and surface properties. Such bearings have high tolerances for boundary and thin-film lubrication conditions, and thus can be used under higher loads than can any of the bimetal systems. Although more costly than the corresponding steel-back bimetal systems, they are used in some high-volume automotive applications as well as in larger mobile and stationary engines. A highly developed body of mechanical, metallurgical, and chemical manufacturing technology has been established in the plain bearing industry, and this technology permits mass production of precision trimetal bearings without a severe cost penalty.

Bearing Alloys

Sliding bearing alloys can be grouped as follows:

- Tin-base alloys
- Lead-base alloys
- Copper-base alloys
- Aluminum-base alloys
- Silver-base alloys
- Zinc-base alloys
- Additional metallic materials (gray cast irons and cemented carbides)
- Nonmetallic materials (nylon, PTFE, carbon-graphite, wood, rubber, and laminated phenolics)

Tin-Base Alloys

Tin-base bearing materials (babbitts) are alloys of tin, antimony, and copper that contain limited amounts of zinc, aluminum, arsenic, bismuth, and iron. The compositions of tin-base bearing alloys, according to ASTM B 23, SAE, and ISO specifications, are shown in Table 5.

Table 5 Designations and nominal composition of tin-base bearing alloys

Designation[(a)]				Composition, %				Product form	Applications
UNS	**SAE**	**ISO**	**ASTM (B 23)**	**Sn**	**Sb**	**Cu**	**Other**		
L13910	. . .	. . .	Alloy 1	91	4.5	4.5	. . .	Cast on steel or bronze back	Bimetal surface layer
L13870	. . .	. . .	Alloy 11	87	7	6	. . .	Cast on steel; bronze or steel-backed bronze	Bimetal and trimetal surface layer
L13890	Alloy 12	SnSb8Cu4	Alloy 2	89	7.5	3.5	. . .	Cast on steel; bronze or steel-backed bronze	Bimetal and trimetal surface layer
L13840	. . .	. . .	Alloy 3	84	8	8	. . .	Cast on steel or bronze back	Bimetal surface layer
. . .	. . .	SnSb12Cu6Pb	. . .	80	12	6	2 Pb	Cast on steel back	Bimetal surface layer

(a) UNS, unified numbering system; SAE, Society of Automotive Engineers; ISO, International Organization for Standardization; ASTM, American Society for Testing and Materials

The presence of zinc in these bearing metals generally is not favored. Arsenic increases resistance to deformation at all temperatures; zinc has a similar effect at 38 °C (100 °F) but causes little or no change at room temperature. Zinc has a marked effect on the microstructures of some of these alloys. Small quantities of aluminum (even <1%) will modify their microstructures. Bismuth is objectionable because, in combination with tin, it forms a eutectic that melts at 137 °C (279 °F). At temperatures above this eutectic, alloy strength is decreased appreciably.

Bulk mechanical properties of ASTM grades 1 to 3 are shown in Table 6. These properties have some value for initial materials screening comparisons of alloys; but they are not reliable predictors of the performance of thin layers bonded to a strong backing, which is the manner in which tin-base babbitts are usually used in modern bearing practice. Layer thickness effects (Fig. 5) and temperature effects (Fig. 7) are more important practical considerations than the mechanical property differences among the various alloy compositions.

Table 6 Typical mechanical properties of chill cast tin-base bearing alloys

Designation				Compressive yield strength[(a)]				Ultimate tensile strength, 20 °C (68 °F)		Elongation, 20 °C (68 °F), %	Hardness, HB	
				20 °C (68 °F)		**100 °C (212 °F)**						
UNS	**SAE**	**ISO**	**ASTM (B 23)**	**MPa**	**ksi**	**MPa**	**ksi**	**MPa**	**ksi**		**20 °C (68 °F)**	**100 °C (212 °F)**
L13910	. . .	. . .	Alloy 1	30.3	4.40	18.3	2.65	64	9.3	2	17.0	8.0
L13890	Alloy 12	SnSb8Cu4	Alloy 2	42.1	6.10	20.7	3.00	77	11.2	. . .	24.5	12.0
L13840	. . .	. . .	Alloy 3	45.5	6.60	21.7	3.15	69[(b)]	10.0[(b)]	1[(b)]	27.0	14.5

(a) 0.125% offset.

(b) Values are for die-cast alloy specimens.

Compared with most other bearing materials, tin alloys have low resistance to fatigue, but their strength is sufficient to warrant their use under low-load conditions. These alloys are commercially easy to bond and handle and they have excellent antiseizure qualities. Their excellent corrosion-resistant properties make these alloys especially well-suited for bearing applications in compressors, electric motors, and food-processing equipment.

Tin-base alloys vary in microstructure in accordance with their composition. Alloys that contain about 0.5 to 8% Cu and <8% Sb are characterized by a solid-solution matrix in which needles of a copper-rich constituent and fine rounded particles of precipitated SbSn are distributed. The proportion of the copper-rich constituent increases with copper content. SAE 12 (ASTM grade 2) has a structure of this type in which the needles often assume a characteristic hexagonal starlike pattern. Alloys that contain ~0.5 to 8% Cu and >8% Sb exhibit primary cuboids of SbSn, in addition to needles of the copper-rich constituent in the solid-solution matrix. In alloys with ~8% Sb and ~5 to 8% Cu, rapid cooling suppresses formation of the SbSn cuboids. This is particularly true of alloys containing lower percentages of copper.

Lead-Base Alloys

Lead-base bearing materials (lead babbitts) are alloys of lead, tin, antimony, and in many cases, arsenic. Many such alloys have been used for centuries as type metals, and were probably first employed as bearing materials because of the properties they were known to possess. The advantage of arsenic additions have been generally recognized since 1938.

Nominal compositions of the most widely used lead babbitts according to ASTM, SAE, and ISO specifications are listed in Table 7. Additional information on mechanical properties of some of these alloys is given in Table 8.

Table 7 Designations and nominal composition of lead-base bearing alloys

Designation				Composition, %					Product form	Applications
UNS	**SAE**	**ISO**	**ASTM (B 23)**	**Pb**	**Sb**	**Sn**	**As**	**Other**		
L53346	Alloy 13	PbSb10Sn6	Alloy 13	84	10	6	. . .	. . .	Cast on steel or steel-backed bronze	Bimetal and trimetal surface layer
L53585	Alloy 14	PbSb15Sn10	Alloy 7	75	15	10	. . .	. . .	Cast on steel back	Bimetal surface layer
L53620	Alloy 15	PbSb15SnAs	Alloy 15	84	15	1	1	. . .	Cast on steel back	Bimetal surface layer
. . .	. . .	PbSb14Sn9CuAs	. . .	77	14	9	0.5	1 Cu	Cast on steel back	Bimetal surface layer
L53565	. . .	. . .	Alloy 8	80	15	5	0.5	. . .	Cast on steel back	Bimetal surface layer
. . .	. . .	. . .	(a)	87.5	9	3.5	. . .	. . .	Cast on bronze back	Bimetal surface layer

(a) ASTM B 67

Table 8 Typical mechanical properties of chill cast lead-base bearing alloys

Designation				Compressive yield strength(a)				Ultimate tensile strength, 20 °C (68 °F)		Elongation, 20 °C (68 °F), %	Hardness, HB	
				20 °C (68 °F)		**100 °C (212 °F)**						
UNS	**SAE**	**ISO**	**ASTM (B 23)**	**MPa**	**ksi**	**MPa**	**ksi**	**MPa**	**ksi**		**20 °C (68 °F)**	**100 °C (212 °F)**
L53346	Alloy 13	PbSb10Sn6	Alloy 13	22.8	3.30	10.7	1.55	69	10.0	5	19	8.5
L53585	Alloy 14	PbSb15Sn10	Alloy 7	24.5	3.55	11.0	1.60	72	10.5	4	22.5	10.5
L53620	Alloy 15	PbSb15SnAs	Alloy 15	24.8	3.60	14.5	2.10	71	10.4	2	21	13

(a) 0.125% offset

Comments made in the section "Tin-Base Alloys" in this article concerning the significance of bulk mechanical properties of tin-base babbitt alloys apply equally to those of lead-base alloys.

For many years, lead-base bearing alloys were considered to be only low-cost substitutes for tin alloys. However, the two groups of alloys do not differ greatly in antiseizure characteristics, and when lead-base alloys are used with steel backs and in thicknesses <0.75 mm (<0.03 in.), they have fatigue resistance that is equal to, if not better than, that of tin alloys. Bearings of any of these alloys remain serviceable longest when they are ≤0.13 mm (≤0.005 in.) thick (See Fig. 5).

In the absence of arsenic, the microstructures of these alloys comprise cuboid primary crystals of SbSn or of antimony embedded in a ternary mixture of Pb-Sb-SbSn in which lead forms the matrix. The number of these cuboids per unit volume of alloy increases as antimony content increases. If antimony content is >15%, the total amount of the hard constituents increases to such an extent that the alloys become too brittle to be useful as bearing materials.

Arsenic is added to lead babbitts to improve their mechanical properties, particularly at elevated temperatures. All lead babbitts are subject to softening or loss of strength during prolonged exposure to the temperatures (95 to 150 °C, or 200 to 300 °F) at which they serve as bearings in internal combustion engines. Addition of arsenic minimizes such softening. Under suitable casting conditions (see the section "Casting Processes" in this article), the arsenical lead babbitts (for example, SAE 15 [ASTM grade 15]) develop remarkably fine and uniform structures. They also have better fatigue strength than arsenic-free alloys.

Pouring temperature and rate of cooling markedly influence the microstructures and properties of lead alloys, particularly when they are used in the form of heavy liners for railway journals. High pouring temperatures and low cooling rates, which typically result from the use of overly hot mandrels, promote segregation and formation of a coarse structure. A coarse structure may cause brittleness, low compressive strength, and low hardness. Therefore, low pouring temperatures (325 to 345 °C, or 620 to 650 °F) usually are recommended. Because these alloys remain relatively fluid almost to the point of complete solidification (~240 °C, or 465 °F, for most compositions), they are easy to manipulate and can be handled with no great loss of metal from drossing.

Lead-Base Electroplated Overlays. The improvement in fatigue life that can be achieved by decreasing babbitt layer thickness has already been noted. Economically as well as mechanically, it is difficult to consistently achieve very thin uniform babbitt layers bonded to bimetal shells by casting techniques. Therefore, the process of electroplating (see the section "Electroplating Processes" in this article) a thin precision babbitt layer on a very accurately machined bimetal shell was perfected. Specially designed plating racks allow the thickness of the plated babbitt layer to be regulated so accurately that further machining is usually not required.

Electroplated tin alloys were found to be generally inferior to lead alloys, and only lead alloys are in commercial use as electroplated bearing overlays. Table 9 lists the four most commonly used compositions. SAE alloy 192 is the most frequently used. Tin in alloys 191, 192, and 193 and indium in alloy 194 impart corrosion resistance. Tin also increases wear resistance. Both copper and indium enhance fatigue resistance.

Table 9 Designations and nominal composition of lead-base electroplated overlay alloys

Designation		Composition, %			
SAE	**ISO**	**Pb**	**Sn**	**Cu**	**In**
Alloy 191	PbSn10	90	10	. . .	. . .
Alloy 192	PbSn10Cu2	88	10	2	. . .
Alloy 193	. . .	80	18	2	. . .
Alloy 194	PbIn7	93	. . .	. . .	7

When a tin-containing overlay is plated directly onto a copper-lead or bronze surface, the tin has a tendency to migrate to the copper interface, forming a brittle copper-tin intermetallic compound. This decreases the corrosion resistance of the overlay and causes embrittlement along the bond line. To avoid this deterioration, a thin continuous barrier layer, preferably nickel ~1.3 μm (~52 μin.) thick, is plated onto the copper alloy surface just prior to plating of the overlay. In addition to providing better surface behavior, overlays improve fatigue performance of some intermediate layers by preventing cracking in this layer. Plated overlays generally range in thickness from 0.013 to 0.05 mm (0.0005 to 0.002 in.), with fatigue life increasing markedly as overlay thickness decreases. In order to take full advantage of the improved fatigue life achieved with thin overlays, it is necessary to minimize assembly imperfections (such as misalignment) and to maintain close tolerances on machined shafts and bearing bores. Engine components must be thoroughly cleaned before assembly, and adequate air and lubricant filtration must be maintained if the overlay is to survive during the useful life of the bearing. Under adverse wear conditions, however, premature removal of the overlay will not necessarily impair operation of the bearing, because the exposed intermediate bearing alloy layer should continue to function satisfactorily.

Copper-Base Alloys

Copper-base bearing alloys comprise a large family of materials with a wide range of properties. They include commercial bronze, copper-lead alloys, and leaded and unleaded tin bronzes. They are used alone in single-metal bearings, as bearing backs with babbitt surface layers, as bimetal layers bonded to steel backs, and as intermediate layers in steel-backed trimetal bearings (see Tables 2, 3, and 4).

The moderate strength and hardness of pure copper are readily increased by alloying, most commonly with tin (with which copper forms a solid solution). Lead is present in cast copper-base bearing alloys as a nearly pure, discrete phase, because its solid solubility in the matrix is practically nil. The lead phase, which is exposed on the running surface of a bearing, constitutes a site vulnerable to corrosive attack under certain operating conditions.

The antifriction behavior of copper-base bearing alloys improves as lead content increases, although at the same time strength is degraded because of increased interruption of the continuity of the copper alloy matrix by the soft weak lead. Thus, through judicious control of tin content, lead content, and microstructure, a large family of bearing alloys has evolved to suit a wide variety of bearing applications.

Table 10 gives specification numbers and nominal compositions of copper-base bearing alloys, as well as the forms in which the alloys are used and general notations on typical product applications. The information in Table 10 should be used in conjunction with the appropriate portions of Tables 2, 3, and 4 and with the brief descriptions that follow.

Table 10 Designations and nominal composition of copper-base bearing alloys

No.	Designation				Composition, %				Product form	Applications
	UNS	**SAE**	**ISO**	**Other**	**Cu**	**Sn**	**Pb**	**Zn**		
Commercial bronze										
1	C83420	Alloy 795	. . .	. . .	90	0.5	. . .	9.5	Wrought strip	Solid bronze bushings and washers
Unleaded tin bronzes										
2	C52100	. . .	CnSn8P	. . .	92[(a)]	8	. . .	. . .	Wrought strip	Solid bronze bushings and washers
3	C90300	. . .	. . .	. . .	88	8	. . .	4	Cast tubes	Solid bronze bearings
4	C90500	. . .	. . .	. . .	88	10	. . .	2	Cast tubes	Solid bronze bearings
5	C91100	. . .	. . .	. . .	84	16	. . .	. . .	Cast tubes	Solid bronze bearings
6	C91300	. . .	. . .	. . .	81	19	. . .	. . .	Cast tubes	Solid bronze bearings
Low-lead tin bronzes										
7	C92200	. . .	. . .	. . .	88.5	6	1.5	4	Cast tubes	Solid bronze bearings
8	C92300	. . .	. . .	. . .	87	8.5	0.5	4	Cast tubes	Solid bronze bearings
9	C92700	. . .	. . .	. . .	87.5	10	2	0.5	Cast tubes	Solid bronze bearings
Medium-lead tin bronzes										
10	C83520	Alloy 791	. . .	. . .	88	4	4	4	Wrought strip	Solid bronze bushings and washers
11	. . .	. . .	. . .	F32/F62	88	4	4	4	Cast on steel back	Bimetal bushings and washers, trimetal intermediate layer
12	C83600	. . .	CuPb5Sn5Zn5	. . .	85	5	5	5	Cast tubes	Solid bronze bearings, bronze bearing backs
13	C93200	. . .	CuSn7Pb7Zn3	. . .	83	7	7	3	Cast tubes	Solid bronze bearings
14	. . .	Alloy 793	. . .	. . .	88	4	8	. . .	Cast on steel back	Bimetal surface layer
15	. . .	Alloy 793	. . .	. . .	88	4	8	. . .	Sintered on steel back	Bimetal surface layer
16	C93700	. . .	CuPb10Sn10	. . .	80	10	10	. . .	Cast tubes	Solid bronze bearings, bronze bearing backs
17	. . .	Alloy 792	CuPb10Sn10(G)	. . .	80	10	10	. . .	Cast on steel back	Bimetal surface layer, trimetal intermediate layer
18	. . .	Alloy 792	CuPb10Sn10(P)	. . .	80	10	10	. . .	Sintered on steel back	Bimetal surface layer
High-lead tin bronzes										
19	C93800	. . .	. . .	. . .	78	7	15	. . .	Cast tubes	Solid bronze bearings, bronze bearing backs
20	. . .	. . .	. . .	AMS 4825	74	10	16	. . .	Cast on steel back	Bimetal surface layer
21	. . .	Alloy 794	CuPb24Sn4(G)	. . .	73.5	3.5	23	. . .	Cast on steel back	Bimetal surface layer, trimetal intermediate layer
22	. . .	Alloy 794	CuPb24Sn4(P)	. . .	73.5	3.5	23	. . .	Sintered on steel back	Bimetal surface layer, trimetal intermediate layer
23	. . .	. . .	. . .	F112	72.5	2.5	25	. . .	Cast on steel back	Trimetal intermediate layer
24	C94300	. . .	. . .	. . .	70	5	25	. . .	Cast tubes	Solid bronze bearings
Copper-lead alloys										
25	. . .	Alloy 49	CuPb24Sn(G)	. . .	75	1	24	. . .	Cast on steel back	Trimetal intermediate layer
26	. . .	Alloy 49	CuPb24Sn(P)	. . .	75	1	24	. . .	Sintered on steel back	Trimetal intermediate layer
27	. . .	Alloy 48	CuPb30(P)	. . .	70	. . .	30	. . .	Sintered on steel back	Bimetal surface layer, trimetal intermediate layer
28	. . .	Alloy 485	. . .	. . .	48	1	51	. . .	Sintered on steel back, infiltrated with lead	Bimetal surface layer

(a) Also 0.3 P

Commercial Bronze. Lead-free copper alloys are characterized by poor antifriction properties but fairly good load-carrying ability. Wrought commercial bronze strip (SAE 795) with 10% Zn can be readily press formed into cylindrical bushings and thrust washers. Strength can be increased by cold working this inexpensive material.

Unleaded Tin Bronze. The unleaded copper-tin alloys are known as phosphor bronzes because they are deoxidized with phosphorus. They are used principally in cast form as shapes for specific applications, or as rods or tubes from which solid bearings are machined. They have excellent strength and wear resistance, both of which improve with increasing tin content, but poor surface properties. They are used for bridge turntables and trunnions in contact with high-strength steel, and in other slow-moving applications.

Low-Lead Tin Bronzes. The inherently poor machinability of tin bronzes can be improved by adding small amounts of lead. Such additions do not significantly improve surface properties, however, and applications for these alloys are essentially the same as those for unleaded tin bronzes.

Medium-Lead Tin Bronzes. The only wrought strip material in this group of alloys is SAE 791, which is press formed into solid bushings and thrust washers. C83600 is used in cast form as bearing backs in bimetal bearings. SAE 793 is a low-tin, medium-lead alloy that is cast or sintered on a steel back and used as a surface layer for medium-load bimetal bushings. SAE 792 is higher in tin and slightly higher in lead; it is cast or sintered on a steel back and used for heavy-duty applications such as wrist pin bushings and heavy-duty thrust surfaces.

High-Lead Tin Bronzes. These contain medium-to-high amounts of tin, and relatively high lead contents to markedly improve antifriction characteristics. SAE 794, widely used in bushings for rotating loads, has the same bronze matrix composition as SAE 793 (4.5% Sn) but three times as much free lead. It is cast or sintered on a steel back and used for somewhat higher speeds and lower loads than alloy 793. The bronze matrix of SAE 794 is much stronger than that of a plain 75-25 copper-lead alloy. Alloy 794 can be used as the intermediate layer with a plated overlay in heavy-duty trimetal bearing applications such as main and connecting-rod bearings in diesel truck engines. This construction provides the highest load-carrying ability available in copper alloy trimetal bearings.

Copper-Lead Alloys. These are used extensively in automotive, aircraft, and general engineering applications. These alloys are cast or sintered to a steel backing strip from which parts are blanked and formed into full-round or half-round shapes depending on final application. Copper-lead alloys continuously cast on steel strip typically consist of copper dendrites perpendicular and securely anchored to the steel back, with an interdendritic lead phase. In contrast, sintered copper-lead alloys of similar composition are composed of more equiaxed copper grains with an intergranular lead phase.

High-lead alloy SAE 48 can be used bare on steel or cast iron journals. Tin content in this alloy is restricted to a minimum value to maintain a soft copper matrix, which together with the high lead content improves the antifriction/antiseizure properties of the alloy. Bare bimetal copper-lead bearings are used infrequently today because the lead phase, present as nearly pure lead, is susceptible to attack by corrosive products that can form in the crankcase lubricant during extended oil-change periods. Therefore, most copper-base alloys with lead contents >20%, including both SAE alloy 48 and alloy 49, are now used with plated overlays in trimetal bearings for automotive and diesel engines.

SAE 485 is a special sintered and infiltrated composite material, produced by methods described in the section "Powder Metallurgy Processes" in this article. By these methods, it is possible to combine a very strong continuous copper alloy matrix structure with a very high lead content, and to alloy the lead-rich constituent with sufficient tin to make it resistant to corrosion. SAE 485 is used principally for bushing and bearing applications that involve alignment, shaft surface finish, or unusual dirt contamination problems.

Mechanical Properties of Copper-Base Bearing Alloys. Table 11 shows the ranges of mechanical strength properties that are exhibited by copper-base bearing alloys, according to alloy families and forms as listed in Table 10. Indentation hardness tests provide the most generally useful indications of behavior under compressive loads, and are the only standard strength tests that are applicable to all of the alloy forms. Conventional tensile and compression tests can be performed only on solid alloy bodies, which represent a relatively small fraction of total copper-base bearing alloy applications.

Table 11 Typical room-temperature mechanical properties of copper-base bearing alloys

Alloy family	Product form	Compressive yield strength[(a)]		Ultimate tensile strength		Hardness, HB
		MPa	**ksi**	**MPa**	**ksi**	
Commercial bronze	Wrought strip	. . .	. . .	310-440	45-64	78-115
Unleaded tin bronzes	Wrought strip	. . .	. . .	400-580	58-84	80-160
	Cast tubes	90-125	13-18	240-310	35-45	70-170
Low-lead tin bronzes	Cast tubes	. . .	. . .	275-290	40-42	65-77
Medium-lead tin bronzes	Wrought strip	. . .	. . .	310-440	45-64	78-115
	Cast tubes	90-100	13-14	240-255	35-37	60-65
	Steel backed	. . .	. . .	. . .	. . .	50-130
High-lead tin bronzes	Cast tubes	75-85	11-12	185-210	27-30	48-55
	Steel backed	. . .	. . .	. . .	. . .	55-90
Copper-lead alloys	Steel backed	. . .	. . .	. . .	. . .	30-80

(a) 0.1% offset

Test information of this kind is helpful in the material selection process as a supplement to information generated in dynamic rig tests and in actual service. Except in certain solid-alloy bearings and bushing applications, alloy strength and hardness values are rarely stated as absolute specification requirements.

Aluminum-Base Alloys

Successful commercial use of aluminum alloys in plain bearings dates back to about 1940, when low-tin aluminum alloy castings were introduced to replace solid bronze bearings for heavy machinery. Production of steel-backed strip materials by roll bonding (see the section "Roll Bonding Processes" in this article) became commercially successful about 1950, permitting the development of practical bimetal and trimetal bearing material systems using aluminum alloys in place of babbitts and copper alloys.

The ready availability of aluminum and its relatively stable cost have provided an incentive for continuing development of its use in plain bearings. Aluminum single-metal, bimetal, and trimetal systems can now be used in the same load ranges as babbitts, copper-lead alloys, and high-lead tin bronzes. Moreover, the outstanding corrosion resistance of aluminum has become an increasingly important consideration in recent years, and has led to widespread use of aluminum alloy materials (in the place of copper-lead alloys and leaded bronzes) in automotive engine bearings.

Designations and Compositions. Alloy designations and nominal compositions of the aluminum-base bearing alloys in most extensive commercial use are listed in Table 12. In these alloys, additions of silicon, copper, nickel, magnesium, and zinc function to strengthen the aluminum through solid-solution and precipitation mechanisms. Fatigue resistance and the opposing properties of conformability and embeddability are largely controlled by these elements and by the use of appropriate heat treatments. Tin and lead are instrumental in upgrading the inherently poor compatibility of aluminum. Cadmium is also used as an alloy addition for this reason. Silicon has a beneficial effect on compatibility in addition to its moderate strengthening effect. Although not well understood theoretically, this compatibility-enhancing mechanism is of considerable practical value. Silicon is used effectively in many alloys for this reason (usually in conjunction with tin, lead, or cadmium).

Table 12 Designations and nominal composition of aluminum-base bearing alloys

Designation					Composition, %							Product form	Applications
No.	**UNS**	**SAE**	**ISO**	**Other**	**Al**	**Si**	**Cu**	**Ni**	**Mg**	**Sn**	**Other**		
High-tin alloys													
1	A08081	Alloy 783	AlSn20Cu	. . .	79	. . .	1	. . .	. . .	20	. . .	Wrought strip, bonded to steel back	Bimetal surface layer
2	. . .	Alloy 786	. . .	. . .	59.5	. . .	0.5	. . .	. . .	40	. . .	Wrought strip, bonded to steel back	Bimetal surface layer
High-lead alloys													
3	. . .	Alloy 787	. . .	F-85	85	4	1	. . .	. . .	1.5	8.5 Pb	Powder rolled strip, bonded to steel back	Bimetal surface layer
4	. . .	Alloy 787	. . .	Al-6	88.5	4	0.5	. . .	. . .	1	6 Pb	Wrought strip, bonded to steel back	Bimetal surface layer
Intermediate-tin alloys													
5	. . .	Alloy 788	. . .	SA-151	82.5	3	1	. . .	. . .	12	1.5 Pb 0.2 Cr	Wrought strip, bonded to steel back	Bimetal surface layer
6	. . .	Alloy 788	. . .	AS-124	82	4	2	. . .	. . .	12	. . .	Wrought strip, bonded to steel back	Bimetal surface layer
7	. . .	Alloy 788	. . .	A-17-X	83	2.5	0.7	. . .	. . .	12	2 Pb 0.2 Sb	Wrought strip, bonded to steel back	Bimetal surface layer
8	. . .	Alloy 788	. . .	FA-130	81	3	0.7	. . .	. . .	13	2 Pb 0.2 Sb 0.2 Sr	Wrought strip, bonded to steel back	Bimetal surface layer
Low-tin alloys													
9	. . .	. . .	AlSn6CuNi	. . .	91.5	. . .	1	1	. . .	6.5	. . .	Cast tubes	Solid aluminum alloy bearings
10	A08500	. . .	. . .	. . .	90.5	. . .	1	1	1	6.5	. . .	Cast tubes	Solid aluminum alloy bearings
11	A08510	. . .	. . .	. . .	89.5	2.5	1	0.5	. . .	6.5	. . .	Cast tubes	Solid aluminum alloy bearings
12	A08520	. . .	. . .	. . .	89.5	. . .	2	1.2	1	6.5	. . .	Cast tubes	Solid aluminum alloy bearings
13	A08280	Alloy 780	. . .	. . .	90.5	1.5	1	0.5	. . .	6.5	. . .	Wrought strip and plate	Solid aluminum alloy bearings
14	. . .	Alloy 770	AlSn6CuNi	. . .	91.5	. . .	1	1	. . .	6.5	. . .	Wrought strip, bonded to steel back	Bimetal surface layer, trimetal intermediate layer
15	A08280	Alloy 780	. . .	. . .	90.5	1.5	1	0.5	. . .	6.5	. . .	Wrought strip, bonded to steel back	Bimetal surface layer, trimetal intermediate layer
Tin-free alloys													
16	A04002	Alloy 781	AlSi4Cd	. . .	95	4	0.1	. . .	0.1	. . .	1 Cd	Wrought strip, bonded to steel back	Bimetal surface layer, trimetal intermediate layer
17	A04002	Alloy 781	AlSi4Cd	F-154	95	4	0.1	. . .	0.1	. . .	1 Cd	Wrought strip, bonded to steel back, precipitation hardened	Trimetal intermediate layer
18	. . .	Alloy 782	AlCd3CuNi	. . .	95	. . .	1	1	. . .	. . .	3 Cd	Wrought strip, bonded to steel back	Bimetal surface layer, trimetal intermediate layer
19	. . .	Alloy	AlSi11Cu	. . .	88	11	1	. . .	. . .	. . .	. . .	Wrought strip,	Trimetal

		784										bonded to steel back	intermediate layer
20	. . .	Alloy 785	AlZn5Si2CuPb	. . .	91.5	1.5	1	. . .	. . .	. . .	5 Zn 1 Pb	Wrought strip, bonded to steel back	Trimetal intermediate layer

Microstructural Features. The cast low-tin alloys (numbers 9 through 11 in Table 12) all display similar microstructures consisting of equiaxed aluminum grains with $NiAl_3$, free silicon (if present), and free tin precipitated in the grain boundaries. Tin forms a nearly complete envelope around each aluminum grain. The copper and magnesium are mostly or completely in solid solution in the aluminum and are not visible under the microscope. Microstructures of the wrought low-tin, intermediate-tin, and high-tin alloys exhibit the expected effects of rolling and annealing, with the as-cast aluminum grains replaced by new recrystallized grains and the insoluble phases ($NiAl_3$, and silicon) uniformly distributed throughout. The original continuous grain boundary envelope of free tin assumes a completely new configuration, the tin now appearing as somewhat elongated, discontinuous "lakes." This characteristic structure, often termed "reticular," results in much greater ductility than that of the cast alloys.

Microstructures of the lead-aluminum alloys (numbers 3 and 4 in Table 12) exhibit a recrystallized aluminum matrix with a fine uniform dispersion of free silicon. The lead is present as thin stringers or ribbons of the lead-tin constituent, elongated in the rolling direction. During recrystallization, this constituent does not coalesce into lakes as does free tin, and the ribbon-like configuration persists in finished bearings. The effectiveness of the modest lead concentrations in these alloys in imparting surface compatibility probably is related to the favorable orientation of the lead-tin ribbons relative to the bearing surface.

The wrought tin-free alloys (numbers 16 through 20 in Table 12) display very simple microstructures, consisting of a recrystallized aluminum matrix with the soluble strengthening additions (copper, zinc, and magnesium) in solid solution. Insoluble phases ($NiAl_3$, silicon, cadmium, and lead) are present in fine and uniform dispersion.

Mechanical Properties of Aluminum-Base Bearing Alloys. Conventional mechanical properties are somewhat like microstructural features in that they are of more value in predicting the fabrication behavior of aluminum-base bearing alloys than in predicting their bearing performance. With the exception of solid aluminum alloy bearings, in which there is no steel back and where press-fit retention depends entirely on the strength of the aluminum alloy, mechanical properties of finished bearings are rarely specified, usually for control purposes only. Consideration of some of these properties (Table 13) does, however, contribute to an understanding of these alloys as a family of related engineering materials, and of their relationship to the better-known structural aluminum alloys in addition to the copper-base, tin-base, and lead-base bearing alloys discussed previously.

Table 13 Typical room-temperature mechanical properties of aluminum-base bearing alloys

Alloy family	Product form	Compressive yield strength[a]		Ultimate tensile strength		Hardness, HB
		MPa	**ksi**	**MPa**	**ksi**	
High-tin alloys	Steel backed	. . .	. . .	100-130	15-19	25-40
High-lead alloys	Steel backed	. . .	. . .	. . .	. . .	40-50
Intermediate-tin alloys	Steel backed	. . .	. . .	. . .	. . .	50-60
Low-tin alloys	Cast tubes	70-140	10-20	125-220	18-32	45-65
	Wrought plate	80-140	12-20	140-170	20-25	40-55
	Steel backed	. . .	. . .	. . .	. . .	35-45
Tin-free alloys	Steel backed	. . .	. . .	. . .	. . .	35-65

(a) 0.2% offset

Product Applications. The majority of the current commercial applications of aluminum-base bearing alloys involve steel-backed bimetal or steel-backed trimetal bearings. To determine the most cost-effective aluminum material for any specific application, consideration should be given to the economic advantages of bimetal versus trimetal systems. The higher cost of the high-tin and high-lead alloys usually is offset by eliminating the cost of the lead alloy overlay plate. The cost effectiveness of the aluminum bimetal materials is clearly demonstrated by the fact that approximately 75% of the passenger car engines built in the United States use high-lead aluminum alloy bimetals for main and connecting-rod

bearings. In Europe and Japan, intermediate and high-tin aluminum bimetals similarly dominate passenger car engine bearing markets.

If the higher load capacity of a trimetal material is required, it then becomes important to select an aluminum liner alloy that provides adequate but not excessive strength, so that conformability and embeddability are not sacrificed unnecessarily. The tin-free alloy group (alloys 16 to 20 in Table 12) offers a wide range of strength properties, and the most economical choice usually is found in this group.

Silver-Base Alloys

Use of silver in bearings is largely confined to unalloyed silver (AMS 4815) electroplated on steel shells, which then are machined to very close dimensional tolerances and finally precision plated to size with a tin overlay of soft metal. The overlay may be lead-tin, lead-tin-copper, lead-indium, or in some cases, pure lead. As a bearing material, plated silver is invariably used with an overlay. Silver on steel with an overlay is regarded as the ultimate fatigue-resistant bearing material.

Silver was widely used during and after World War II in aircraft applications, where its high cost could be justified. With the phasing out of piston engines, however, the use of silver in bearings has greatly declined. Current applications are specialized, chiefly in the aircraft and locomotive industries. In view of the rapidly rising cost of silver, any increase in demand for this material would stimulate the search for a comparable less-expensive substitute.

Zinc-Base Alloys

The zinc-base alloys that have been used successfully for machinery bearings are standard zinc foundry alloys of the zinc-aluminum-copper-magnesium high-performance type. Tubular shapes made by conventional sand, permanent mold, and pressure die-casting methods are machined into bearings in the same way that solid bronze bearings are made. Most applications have been direct substitutions for solid bronze bearings; the substitutions are made primarily to reduce costs. Alloy designations, nominal compositions, and typical mechanical properties are shown in Tables 14 and 15 for the two predominant alloys in the United States.

Table 14 Designations and nominal composition of zinc-base bearing alloys

Designation		Composition, %			
UNS	**Trade**	**Zn**	**Al**	**Cu**	**Mg**
Z35631	ZA-12	88	11	1	0.025
Z35831	ZA-27	71	27	2	0.015

Table 15 Typical room-temperature mechanical properties of zinc-base bearing alloy

Designation		Casting method	Compressive yield strength[(a)]		Ultimate tensile strength		Elongation, %	Hardness, HB
UNS	**Trade**		**MPa**	**ksi**	**MPa**	**ksi**		
Z35630	ZA-12	Sand cast	230	33	299	43	1.5	94
		Permanent mold cast	234	34	328	48	2.2	110
		Die cast	269	39	404	59	5	125
Z35840	ZA-27	Sand cast	330	48	421	61	4.5	90

(a) 0.1% offset

The high compressive strength and hardness values of these materials suggest greater load capacities than those of solid bronze and solid aluminum bearing materials (Tables 2 and 11). This is not the case in practice, however, largely because of the high rate at which the zinc alloys soften with increasing temperature. Maximum recommended running temperatures of 95 to 120 °C (205 to 250 °F) for the zinc-base alloys are approximately 100 °C (180 °F) below the temperature limits for copper-base and aluminum-base bearing alloys.

Microstructurally, these alloys display a eutectic or peritectic aluminum-zinc matrix, surrounding zinc-rich or aluminum-rich primary dendrites. Copper is in solid solution. Grain size varies greatly with casting method--from coarse in sand castings to finest in pressure die castings. Some experimental evidence associates the coarse sand cast structures with superior bearing wear resistance.

Because of their low cost, the zinc-base alloys will probably continue to replace copper-base alloys in certain bearing applications in construction, earth-moving, mining, and mill machinery markets. However, technical limitations with respect to high-temperature strength and corrosion resistance will prevent any massive movement away from bronzes and into zinc alloys.

Additional Metallic Bearing Materials

Additional commercially available metallic bearing materials include gray cast irons and cemented carbides.

Gray cast irons are standard materials for certain applications involving friction and wear (for example, brake drums, piston rings, cylinder liners, and gears). Cast irons perform well in such applications, and thus should be given consideration as bearing materials. Gray iron bearings have proved successful in refrigeration compressors where bearing pressures are <4500 kPa (<650 psi) for main bearings and <5500 kPa (<800 psi) for connecting rod bearings. Normally, the journals in refrigeration compressors are made either of steel (carburized and hardened to 55 to 60 HRC) or of pearlitic malleable or ductile iron (hardened to 44 to 48 HRC and having a surface finish of $\leq$0.3 μm, or 12 μin., root-mean-square, R_q. Because of occasional dilution of the oil with liquid refrigerant and heavy foaming of the oil, lubrication may become marginal for short periods of time. Fine-grain iron with uniformly distributed No. 6 (or finer) graphite flakes usually performs well during these periods. The bearings are often phosphate coated to improve their seizure resistance. This type of coating also creates a spongelike surface that promotes retention of oil.

For good wear resistance, gray cast iron should be pearlitic with randomly distributed graphite flakes. Cast irons have been heat treated to obtain martensitic structures for use as cylinder liners, but the benefits of such heat treatment have not been economically justifiable. Hardened cast iron has been used successfully as a material for the ways on machine tools.

Cemented Carbides. Extremely hard materials, including cemented tungsten carbides, titanium carbides, and other combinations have been used successfully for various specialized bearing and seal applications. In terms of the bearing performance characteristics listed in Table 1 these materials exhibit essentially zero conformability and embeddability, but rank high in strength, hardness, corrosion resistance, and compatibility. Cemented carbides have been of greatest interest in high-temperature aerospace applications, but have also been used to advantage in certain machinery and machine tool applications.

Nonmetallic Bearing Materials

Nonmetallic bearing materials are widely used for a variety of applications. They have many inherent advantages over metals, including better corrosion resistance, lighter weight, better resistance to mechanical shock, and the ability to function with very marginal lubrication or with no lubricant present at all. The major disadvantages of most nonmetallics are their high coefficients of thermal expansion and their low thermal conductivity characteristics. For many years, carbon-graphites, wood, rubber, and laminated phenolics dominated the field of nonmetallic bearing materials. In the early 1940s, development of nylon and polytetrafluoroethylene (PTFE, or Teflon) gave engineering designers two new nonmetallics with very unique characteristics, particularly the ability to operate dry.

A wide variety of polymer composites is now being used very successfully in bearing applications. The addition of fiber reinforcements and fillers such as solid lubricants and metal powders to the resin matrix can significantly improve the physical, thermal, and tribological properties of these plastics.

Casting Processes

Single-Metal Systems. Except for porous metal oil-impregnated bushings, all the single-metal systems listed in Table 2 are commercially produced by casting, either with or without subsequent mechanical working. Plate, strip, and sheet forms of commercial bronze, of low-lead and lead-free tin bronzes, and of aluminum-tin alloys are initially cast as ingots, slabs, or bars by static and continuous casting methods similar to those used for other brass and aluminum mill products. Subsequent rolling and annealing operations are also similar to those used for conventional mill products. Because of the

extreme hot shortness of leaded tin bronzes and aluminum-tin alloys, these alloys must be rolled either cold or at only slightly elevated temperatures, with frequent intermediate annealing.

The recrystallized wrought structures of bronze and aluminum-tin bearing alloys are substantially different from the initial cast structures, with respect to the configurations of the copper and aluminum phases and of the free-lead and free-tin phases. The improvements in ductility and forming characteristics that result from these structural changes are of great importance in subsequent bearing manufacturing operations. Bearing performance properties are not strongly affected by these changes. Both the as-cast and wrought forms of these alloys are in commercial use and are equally acceptable in bearing applications.

Tubular and cylindrical bronze, zinc, and aluminum-tin alloy shapes are produced by static, centrifugal, and continuous casting methods, and subsequently are machined into bearings. The high-lead bronzes are used only in the as-cast condition because of their low ductility and extreme hot shortness, which preclude any substantial amount of plastic deformation of cast shapes. Cast aluminum-tin alloy tubes can withstand a limited amount of cold work, however, and in some instances cold compression of 4 to 5% is employed to increase yield strength and improve press-fit retention in the finished bearings.

Bimetal Systems. Specialized casting methods are widely employed for producing bimetal bearing materials in both tubular and flat strip forms. Except for aluminum alloy systems (Table 3, classes 3 and 8), all bimetal systems in commercial use can, at least in principle, be produced by casting methods, and systems that incorporate tin and lead babbitt liners >0.1 mm (>0.004 in.) thick are universally produced by casting.

Babbitt Centrifugal Casting. Short tubular steel and bronze shapes (bearing shells) are commonly lined with tin or lead alloys by various forms of centrifugal casting. In these processes, a machined steel or bronze shell is first pre-heated and coated by immersion in molten tin or tin alloy. The prepared shell is then placed in a lathelike "spinner" and rotated at a controlled speed about its axis. Molten babbitt is admitted through one end and is uniformly distributed around the inside wall of the shell by centrifugal action. The molten layer then is cooled and solidified by spraying water against the outside of the rotating bearing shell. When properly controlled, these processes produce fine-grain liner layers of reasonably uniform thickness, completely bonded to the steel or bronze bearing-back material. Centrifugal casting methods are especially well suited to large-diameter thickwall bearings, which are made in relatively small quantities, and to full-round seamless bearings, which cannot be fabricated from flat strip.

Bronze Centrifugal Casting. Leaded tin bronzes also can be applied to the inner walls of steel shells by centrifugal casting. Various methods of shell preparation are employed, including both molten-salt and controlled-atmosphere pre-heating. Oxidation must be entirely absent from the inner wall of the steel shell for complete metallurgical bonding. Centrifugal casting of bronzes is most successful with alloys containing >3% Sn and $\leq$20% Pb. Outside this composition range, leaded tin bronze and copper-lead alloys are sensitive to lead segregation and consequent nonuniform "centrifuged" microstructures. Within these composition limits and under well-controlled process conditions, mechanically sound well-bonded bronze layers with reasonably uniform microstructures can be produced.

Bronze Gravity Casting. All copper-lead alloys and leaded bronzes containing $\leq$35% Pb can be successfully cast in and bonded to steel shells by gravity casting methods, in which centrifugal forces are not a factor. In these processes, a core usually is used to form an annular space inside the shell, into which molten bronze or copper-lead alloy is poured. Several different processes of this kind are in commercial use, utilizing a variety of preheating methods, core materials, pouring methods, and quenching procedures.

As in centrifugal casting, absence of oxides on the inner shell wall is necessary for complete bonding of the alloy layer to the steel back. Liner microstructures produced by gravity shell casting methods generally are more uniform than those obtained by centrifugal casting. For low-tin and high-lead compositions, gravity casting is preferred because of the absence of centrifuging effects on the solidifying alloy.

Babbitt Strip Casting. Steel-backed tin alloy and lead alloy bearing strip materials are commonly produced by continuous casting in specially designed process lines in which separate cleaning, etching, hot tinning, liner-alloy casting, and quenching operations are carried out continuously on a moving steel bearing-back strip. One or more in-line machining operations may also be incorporated so that the strip emerges with a closely controlled thickness, suitable for bearing fabrication.

Bronze Strip and Slab Casting. The oldest commercial processes for producing steel-back copper-lead and leaded bronze bearing strip also utilize continuous casting on a moving steel strip. Steel preheating, alloy casting, and quenching operations are performed under a strongly reducing atmosphere to ensure freedom from oxidation. Some in-line machining also can be done, but the cast strip usually is machined in a separate line for close control of thickness. Additional cold rolling and annealing operations are also employed--particularly with the low and medium lead-tin bronze alloys, in which recrystallized structures are frequently preferred for their superior fabrication properties.

Strip casting of copper alloys is a difficult technology that requires close process control, a high level of operator skill, and relatively expensive special-purpose equipment. It is used by only a few bearing manufacturers, but with considerable commercial success. It is employed not only for thin-gage coiled materials but also for heavy-gage slabs with steel thicknesses as great as 15 mm (0.60 in.).

Trimetal Systems. Trimetal materials with relatively thick surface layers (Table 4, classes 1, 2, and 3) are used mostly in large bearings. These bearings are produced in relatively low volumes from steel shells initially lined by casting with copper-lead alloys or bronze. After intermediate machining to remove excess liner alloy, such shells are commonly relined with tin or lead babbitt by centrifugal casting. The methods used are essentially the same as those for casting in bare steel or solid bronze shells.

Powder Metallurgy Processes

Single-Metal Systems. The only commercial use of powder metallurgy (P/M) methods for making single-metal bearing materials is in the fabrication of copper-base and iron-base porous metal bushings, which are subsequently impregnated with oil. The methods used are similar to those for making structural P/M shapes (that is, pressing in a closed die and sintering under a reducing atmosphere). Bars, tubes, and finished parts are made in this way. Post-sinter coining and repressing operations are frequently used to control final dimensions of finished parts.

Bimetal and Trimetal Systems. No powder metallurgy processes that use lead-base or tin-base bearing alloys are in commercial use nor are there at present any commercially developed processes for lining bearing shells by means of P/M methods. In the manufacture of steel-back copper-lead alloy and leaded bronze strip, however, P/M methods are employed more extensively than any other method.

Continuous Sintering Process. A wide variety of steel-back copper alloy materials, including counterparts of all of the cast copper-lead and leaded bronze bearing alloys (Table 3, class 4), can be produced by continuous sintering on a steel backing strip. In these processes, prealloyed (PA) powder particles are spread uniformly onto moving steel strip. As the strip passes through a furnace under a reducing atmosphere, the particles become sintered together, forming an open grid bonded to the steel strip. After cooling, this bimetal is rolled to densify the liner alloy and then resintered to develop complete interparticle and alloy/steel bonds. After resintering, the strip material may receive further rolling--to attain finish stock size, and sometimes to strain harden the alloy liner for increased strength.

Strip sintering technology makes possible the production of steel-core "sandwich" material, which is especially suitable for applications requiring two bearing surfaces (such as in some thrust washers). In this instance, powder spreading, sintering, cooling, and rolling are repeated on the opposite side, after which the strip is finally resintered. Sintered strip for most automotive and truck bearing applications is processed in coils $\leq$5mm ($\leq$0.2 in.) in overall thickness. Thick-wall materials with steel layers up to about 16 mm ($\frac{5}{8}$ in.) thick also can be processed into flat slab lengths.

Impregnation and Infiltration. Both bimetal and trimetal bearing materials also can be made by impregnation or infiltration of a lower-melting lead alloy into a layer of sintered copper alloy powder. In impregnation, a bilayer strip made from PA copper-lead alloy or leaded bronze powder is immersed in a bath of molten lead-tin alloy heated above the melting point of lead. During immersion, some of the lead at the surface of the strip is replaced by the lead-tin alloy. In infiltration, the copper alloy powder layer is free-sintered and not compacted after sintering. The open-grid sintered layer is then infiltrated with material having a lower melting temperature than that of the grid alloy.

The infiltrant is usually molten lead or a lead alloy but it can be a nonmetallic material such as PTFE, which can be introduced in paste or slurry form. A very useful class of self-lubricating trilayer structures is produced commercially in this way; the PTFE-base infiltrant also forms a thin low shear strength surface film in these structures.

Powder Rolling. One very useful application of direct powder rolling that has been developed commercially in the plain bearing industry is production of an aluminum-lead alloy strip for subsequent bonding to a steel back (see third item under class 3 in Table 3). In this method, PA lead-aluminum powder and unalloyed aluminum powder are fed simultaneously in separate streams to a powder rolling mill and continuously compacted into a bilayer aluminum strip. After sintering, this strip is roll bonded to low-carbon steel, with the unalloyed aluminum bonding layer next to the steel. This steel-back strip is used as a bimetal material for bearing applications where the unit loading is beyond the capability of tin or lead babbitt bimetal material.

Roll Bonding Processes

Virtually all commercial manufacturer of bimetal aluminum alloy bearing strip materials (see Table 3, class 3) is currently done by roll bonding the liner alloy to a steel backing strip. Both batch and continuous processes are employed, the latter being favored for economical high-volume processing of lighter-gage material.

In all roll bonding processes, whether batch or continuous, very clean aluminum and steel surfaces are forced together under intense pressure in a rolling mill, so that solid-phase bonding (cold welding) can occur between the two metals at many sites in the interface. Heat, which may be applied simultaneously with pressure in hot rolling and subsequently in postroll annealing, serves to develop complete diffusion bonding from the initial weld sites and to recrystallize the aluminum alloys so that the final bimetallic strip product exhibits useful liner-alloy ductility and complete bonding.

Tin-aluminum alloys usually are not bonded directly to steel because of undesirable interactions between the free tin constituent and the steel backing. A layer of electrolytic nickel plating on the steel surface is commonly used to alleviate these effects with both low-tin and high-tin alloy compositions.

Another method commonly used with tin-aluminum alloys employs a tin-free aluminum interlayer. This is accomplished by the use of Alclad tin-aluminum alloy strip. The tin-free cladding layer serves as the bonding surface and is present as a distinct bond interlayer in the finished bimetal strip.

Direct roll bonding to steel is most commonly employed with tin-free aluminum alloys (fifth item under class 3, Table 3) and with lead-aluminum strip materials.

Electroplating Processes

Plated Overlays. Lead alloy surface layers (overlays) whose thickness must be limited to <0.05 mm (<0.002 in.) (Table 4, classes 4 to 12) are most commonly produced by electroplating the lead alloy on bimetallic bearings that have previously been finish machined. Specially designed plating racks are used to ensure uniform distribution of plating current over the bearing surface. With close control of current, critical dimensional tolerances often can be maintained so precisely that no machining of the electrodeposited alloy surface is required. One manufacturer has commercialized a process in which the lead alloy electroplating is applied continuously to precision-rolled bimetal strip. In this process, all forming and machining operations are done after electroplating.

Electroplated lead babbitts comprise booth binary lead-tin and ternary lead-tin-copper compositions, all of which are commercially codeposited from fluoborate electrolytes. To ensure against bond and plate defects, extreme care is exercised in preparing the basis metal.

In addition to cleaning and etching treatments, preplating basis-metal preparation usually includes deposition of one more very thin metallic interlayers. A thin layer of nickel is most frequently used over copper-lead alloys and bronzes to prevent diffusion of tin from the plated surface layer into the copper basis metal. Copper is most often used over aluminum alloys to ensure complete adhesion of the plated lead alloy layer, and nickel sometimes is plated over the copper to prevent diffusion of tin from the lead alloy layer into the copper layer.

Binary lead-indium alloy overlays are also used with copper-lead and leaded bronze intermediate layers. These alloys are produced by electroplating separate layers of pure lead and pure indium and subsequently diffusing the indium into the lead in a low-temperature heat treatment operation. In this case, no diffusion barrier is required between the overlay and the intermediate alloy layer.

Plated Silver Intermediate Layers. Pure silver and silver-lead alloy bearing liners are applied to steel shells by electrodeposition from cyanide plating baths. Final machining usually is done after plating, leaving a substantially thick

layer (typically 0.25 to 0.38 mm, or 0.010 to 0.015 in.) of bonded silver liner material. Although as-plated thickness tolerances are not critical, special racking and masking techniques are employed to restrict plating to the surfaces where it is required and to eliminate local concentrations of high current density. If the structure of the plated layer is to be uniform, and the bond strength of the liner uniformly high, the steel basis metal must be prepared very carefully and plating-bath compositions and cleanness must be properly controlled. Although the basic principles involved in silver plating of bearing liners are the same as for decorative silver plating, the unusually thick deposits involved (normally >0.50 mm, or 0.020 in.) and the extremely high quality requirements for bond and plated-metal soundness have led to development of several unique operating and control practices.

Bearing Material Selection

It must be emphasized that selection of a bearing material system for a specific application and of a mechanical design for the bearing itself are closely interrelated processes. Neither process is entirely straightforward, neither can be approached independently, and both require a good understanding of other interacting components of the machine system.

Although this article considers the principles involved in bearing operation, it has not attempted to present a detailed discussion of mechanical design factors. The reader should therefore not expect to make final decisions on materials for specific applications on the basis of this text alone.

Most manufacturers of plain bearings have experienced engineering staff personnel available to aid potential users with both mechanical design and material selection. Because of the wide material selection offered by most of these experienced specialized producers and their background of experience in practical applications, full advantage should be taken of the engineering services such sources can provide.

Selected References

- Bearing and Bushing Alloys, SAE J459c, SAE Information Report, *SAE Handbook 1990,* Part 1, Society of Automotive Engineers, 1990
- Bearing and Bushing Alloys, SAE J460e, SAE Information Report, *SAE Handbook 1990,* Part 1, Society of Automotive Engineers, 1990
- E.R. Booser, Bearing Materials and Properties, *Mach. Des.,* 10 Mar 1966, p 22-28
- K.G. Budinski, *Surface Engineering for Wear Resistance,* Prentice-Hall, 1988, p 15-42
- *Bushing and Thrust Washer Design Manual,* Clevite Engine Parts Div., Gould Inc., 1973, p 23-32
- T. Calayag and D. Ferres, "High Performance, High-Aluminum Zinc Alloys for Low-Speed Bearings and Bushings," Technical Paper Series Paper No. 820643, Society of Automotive Engineers, 1982
- "Custom Plain Bearing Products," Engine Parts Div., Gould Inc., 1986
- G.J. Davies, G.S. Senior, and O. Beaurepaire, "The Development and Application of Polymer Bearings," Technical Information Paper No. 3, Glacier-Vandervell, Inc. (Great Britain), 1990
- A.O. DeHart, Basic Bearing Types, *Mach. Des.,* 10 Mar 1966, p 15-21
- "Fluid Film Bearing Products," Bushings and Bearings Div., JPI Transportation Products, Inc., 1990
- M.L. MacKay, L.J. Cawley, and G.R. Kingsbury, "A New Aluminum-Lead Bearing Material for Automotive Engine Service," Technical Paper Series Paper No. 760113, Society of Automotive Engineers, 1976
- J. Masounave and G. Huard, Comparison between Continuously Cast and Sand Cast Zinc-Aluminum Alloys Used in Bearing Applications, *Wear Resistance of Metals and Alloys,* conference proceedings, ASM International, 1988, p 65-71
- I.D. Massey, N.A. MacQuarrie, D.R. Eastham, "Development of Crankshaft Bearing Materials for Highly Loaded Applications," Technical Information Paper No. 2, Glacier-Vandervell, Inc. (Great Britain), 1990
- S. Mohan, V. Agarwala, and S. Ray, Wear Characteristics of Rheocast and Stircast Al-Pb Metal-Metal Composites, *Tribology of Composite Materials,* conference proceedings, ASM International, 1990, p 189-193

- L.J. Pesek and W.A. Weinkamer, Strip-Type Bearings, *Mach. Des.,* 10 Mar 1966, p 35-39
- "Plain Bearings--Copper Alloys--Part 1: Cast Copper Alloys for Solid and Multilayer Plain Bearings," International Standard 4382/1, International Organization for Standardization, 1982
- "Plain Bearings--Copper Alloys--Part 2: Wrought Copper Alloys for Solid Plain Bearings," International Standard 4382/2, International Organization for Standardization, 1982
- "Plain Bearings--Lead and Tin Casting Alloys for Multilayer Plain Bearings," International Standard 4381, International Organization for Standardization, 1981
- "Plain Bearings--Metallic Multilayer Materials for Thin-Walled Plain Bearings," International Standard 4383, International Organization for Standardization, 1981
- G.C. Pratt and C.A. Perkins, "Aluminum Based Crankshaft Bearings for the High Speed Diesel Engine," Technical Paper Series Paper No. 810199, Society of Automotive Engineers, 1981
- G.C. Pratt and W.J. Whitney, "Progress with Aluminum-Lead Crankshaft Bearing Alloys," Technical Paper Series Paper No. 890552, Society of Automotive Engineers, 1989
- A.E. Roach and C.L. Goodzeit, Why Bearings Seize, *General Motors Engineering Journal,* Sept-Oct 1955
- K. Sakamoto, Y. Ogita, Y. Sato, and T. Tanaka, "Development of New Aluminum-Zinc-Silicon Bearings for Heavy Load Applications in Uprated Engines," Technical Paper Series No. 900124, Society of Automotive Engineers, 1990
- "Standard Designations for Copper and Copper Alloys," Application Data Sheet, Copper Development Association Inc., 1990
- "Standards Handbook, Part 2--Alloy Data, Wrought Copper and Copper Alloy Mill Products," Copper Development Association, 1985
- "Standards Handbook, Cast Copper and Copper Alloy Products, Part 7--Alloy Data, Copper Development Association, 1978
- "Standard Specification for Car and Tender Journal Bearings, Lined," B 67-88, *Annual Book of ASTM Standards,* ASTM, 1988
- "Standard Specification for White Metal Bearing Alloys," B 23-83, *Annual Book of ASTM Standards,* ASTM, 1984
- D.F. Wilcox and E.R. Booser, *Bearing Design and Application,* McGraw-Hill, 1957, p 367-391

Friction and Wear of Hardfacing Alloys

Paul Crook, Haynes International, Inc.; Howard N. Farmer, Consultant

Introduction

HARDFACING may be broadly defined as the application of a wear-resistant material, in depth, to the vulnerable (or worn) surfaces of a component by a weld overlay or thermal spray process. This discussion deals with the weld overlay materials used to resist wear; the thermal spray materials are covered in the article "Thermal Spray Coatings" in this Volume.

The weld overlay materials fall into five categories:

- Build-up alloys
- Metal-to-metal wear alloys
- Metal-to-earth abrasion alloys
- Tungsten carbides (for extreme earth sliding and cutting wear)

- Nonferrous alloys

Out of a total hardfacing alloy market of about 18 × 10^6 kg (40 × 10^6) per annum, it has been estimated that the first four categories, which are iron-based, account for about 90% by weight, but only about 75% by dollar volume.

The build-up alloys (which are essentially highly alloyed steels) are not designed to resist wear but to return a worn part back to, or near, its original dimensions and to provide adequate support for subsequent layers of true hardfacing materials. However, these alloys are sometimes used as wear-resistant materials under mild wear conditions. Typical examples of applications where build-up alloys are used for wearing surfaces include:

- Tractor rails
- Railroad rail ends
- Steel mill table rolls
- Large slow-speed gear teeth

The metal-to-metal wear alloys are martensitic air-hardening steels that, with care, can be applied (without cracking) to wearing areas of machinery parts. Typical applications of this alloy family include:

- Undercarriage components of tractors and power shovels
- Steel mill work rolls
- Crane wheels

Most of the materials in the final three categories consist of hard particles within a metallic matrix, and, for many, it is the hard constituent that provides resistance to wear. The cobalt-base alloys are an exception in this regard because they exhibit resistance to a wider variety of wear forms, largely by virtue of the deformation and fracture characteristics of the cobalt-rich matrix.

The primary function of the metal-to-earth abrasion alloys and tungsten carbides is abrasion resistance. The metal-to-earth alloys are high-chromium white irons in which chromium carbides are formed during alloy solidification. The tungsten carbides are actually composite materials, and their use involves the transfer of discrete tungsten carbide particles (which in the welding consumable forms are encased in a steel tube) across the welding are and into the molten weld pool, where they are subsequently "frozen" into the overlay structure by the matrix formed from the melting of the steel tube.

The metal-to-earth alloys possess resistance to sliding and crushing (that is, low- and high-stress) abrasion on a moderate scale, whereas the tungsten carbides are intended for use under extreme sliding and cutting conditions. Typical applications for these material types include:

- Shovel teeth
- Rock crusher parts
- Plowshares
- Auger flights

The nonferrous hardfacing alloys are used in environments that are too aggressive for the ferrous hardfacing alloys or where high resistance to a specific type of wear (other than abrasion) is required. For the purpose of this article, they are classified as follows:

- Cobalt-base/carbide type
- Cobalt-and nickel-base/Laves type
- Nickel-base/boride type

- Bronze type

Also included in the discussion of the first of these types are several substitute materials, some of which are iron-base, but share with the cobalt-base alloys resistance to a variety of wear conditions and corrosive media.

Whereas the ferrous hardfacing alloys and tungsten carbide composites find extensive use in agriculture, mining, construction, and the steel industry, the nonferrous alloys are typically used in the chemical processing, power, automotive, and oil industries, all of which require resistance to a hostile environment in addition to resistance to wear. The cobalt-base alloys are especially to deformation and chemical attack at high temperatures (500 °C < $T \leq$ 900 °C, or 930 °F < $T \leq$ 1650 °F) and are used to protect dies and guide rolls in the steel industry. Other common applications of the nonferrous hardfacing alloys include:

- Valve seating surfaces (both control valves and diesel exhaust valves
- Pump parts
- Extrusion screw flights
- Rock bit bearings
- Marine bearings
- Glass molding hardware

The discussion up to this point has been devoted to briefly introducing the five families of weld overlay material. It is now appropriate to distinguish, in a general way, between the individual family members. This is best done by considering that within each family (or category), several levels of abrasion resistance are available. Increasing abrasion resistance, however, is accompanied by increasing brittleness, which can cause cracking during cooling, after welding, or under impact loading in service. In certain situations (for example, seating surfaces in chemical control valves), a material with less than optimum abrasion resistance must be chosen in order to obtain a crack-free overlay or to ensure that the overlay does not crack in service.

This "abrasion level" concept is particularly applicable to the metal-to-earth alloys, the tungsten carbide composites, the cobalt-base/carbide-type alloys, and the nickel-base/boride type alloys. The level of abrasion resistance within these material groups is generally proportional to the carbon, hence carbide, or the boron content.

Although the cracking of weld overlays is of great concern in many applications, it may be desirable (for example, for the relief of residual stresses) for others. In the mining and construction industries, where brittle metal-to-earth abrasion alloys are used, relief of stresses prior to the use of equipment is important, the objective being to reduce the risks of underbead (substrate/overlay interface) cracking during service and the subsequent spalling of the overlay.

Several welding processes are used to apply the hardfacing materials. They range from the traditional (for example, oxyacetylene torch) to the new and sophisticated (for example, plasma-transferred arc, PTA, and synergic metal inert gas, MIG). To accommodate these different overlay processes, the hardfacing materials are available in a variety of forms. The most popular processes, and the forms most commonly associated with each process, are:

Weld overlay process	Consumable form
Oxyacetylene	Bare cast or tubular rod
Shielded metal arc (SMA)	Coated solid or tubular rod (stick electrode)
Gas tungsten arc (TIG)	Bare cast or tubular rod
Gas metal arc (MIG)	Tubular or solid wire
Open arc	Tubular wire (flux cored)
Submerged arc	Tubular or solid wire
Plasma-transferred arc (PTA)	Powder

In choosing the process to be used, the following factors are important:

- Deposition rate
- Level of dilution (the amount of substrate material intermixed with the overlay material during welding)
- Portability of equipment

To overcome the effects of dilution on wear properties, multiple layers are often employed.

Build-Up Alloys

Although designed specifically to bring worn parts back to (or close to) their original dimensions and to provide adequate support for subsequent layers of the high-chromium, white irons, and tungsten carbide composites, the build-up materials are sometimes used to resist combinations of impact and light abrasion (particularly where multiple layers and machinability are required). The build-up alloys fall into two categories: alloy steels and austenitic manganese steels. The alloy steel build-up materials are used with carbon steel and alloy steel substrates; the austenitic manganese steels are used for the joining, repair, and/or protection of manganese steel components.

The high strength of the austenitic manganese steels is the result of a synergism between manganese and carbon (Ref 1). In suppressing the formation of phases other than austenitic, manganese not only increases carbon solubility at lower temperatures but also encourages carbon supersaturation of the structure. Both high inherent strength and a high work hardening rate result from this.

Because the austenitic manganese steels are metastable, problems of carbide embrittlement tend to arise when the alloys are cooled slowly or reheated. Manganese steel components are therefore kept as cool as possible during the build-up (repair) process. Often, the bulk of the part is submerged in water during the welding process.

An overview of the abrasion and impact properties of the build-up alloys relative to the metal-to-metal wear alloys is presented in Table 1. Specific build-up alloy compositions are given in Table 2. In the austenitic manganese steels, chromium is added to increase strength but promotes the formation of embrittling carbides. Molybdenum and nickel, on the other hand, suppress the precipitation of embrittling carbides (molybdenum by modifying the type and geometry of carbides formed, and nickel by increasing the stability of the austenitic structure).

Table 1 Impact resistance and abrasion resistance properties of build-up alloys and metal-to-metal wear alloys

Substrate	Resistance[(a)]			
	Impact		**Abrasion**	
	Min	**Max**	**Min**	**Max**
Build-up weld overlay				
Carbon steel	4.0	4.5	0.5	1.0
Manganese steel	8.5	9.5	2.0	2.5
Metal-to-metal weld overlay				
Easy-machining steel	3.0	3.5	1.5	2.0
General-purpose steel	2.5	3.5	1.0	3.5
High-temperature steel	3.0	3.5	2.0	3.5

(a) On scale from 0 to 10, where 10 is optimum resistance property. Min, minimum; max, maximum

Table 2 Composition, hardness, and abrasion data for build-up alloys and metal-to-metal wear alloys

Alloy	Composition, wt%							Hardness, HRC	Abrasion, volume loss			
									Low-stress[c]		**High-stress[d]**	
	Fe	**Cr**	**C**	**Si**	**Mn**	**Mo**	**Ni**		**mm^3**	**in.^{3} × 10^{-3}**	**mm^3**	**in.^{3} × 10^{-3}**
Build-up weld overlay												
EFe 1[a]	bal	2	0.1	1.0	1	1.5	. . .	37	88	5.4	49	3.0
EFeMn-C[a]	bal	4	0.8	1.3	14	. . .	4	18	65	4.0	57	3.5
EFeMn-Cr[a]	bal	15	0.5	1.3	15	2.0	1	24	93	5.7	46	2.8
Metal-to-metal weld overlay												
EFe2[a]	bal	3	0.2	1.0	1	1.0	1	48	54	3.3	66	4.0
EFe3[a]	bal	6	0.7	1.0	1	1.0	. . .	59	60	3.7	68	4.1
ER420[b]	bal	12	0.3	1.0	2	. . .	. . .	45	84	5.1	62	3.8

(a) Two-layer SMA deposit process.

(b) Two-layer SAW deposit process.

(c) Dry sand/rubber wheel test (ASTM G 65, Procedure B):load 13.6 kg (30 lb); 2000 rev.

(d) Slurry/steel wheel test(ASTM B 611, modified):load 22.7 kg (50 lb); 250 rev

Room-temperature hardness and low- and high-stress abrasion data for the compositions are also listed in Table 2. The low-stress data were generated using the dry sand/rubber wheel (ASTM G 65) test, and the high-stress results were produced using the slurry/steel wheel (ASTM B 611, modified) test. The high-stress test is the more severe, and the volume losses reported are for only 250 rev of the steel wheel (as compared with 2000 rev of the rubber wheel in the low-stress test).

As may be deduced from these data, as-deposited hardness is a poor indicator of low- or high-stress abrasion resistance for these alloys.

Metal-to-Metal Wear Alloys

Alloys in the metal-to-metal category (commonly referred to as machinery hardfacing alloys) are martensitic, air-hardening steels that can be applied several layers thick and can be finish machined (although with some difficulty) if adequate equipment is used. They are recommended for wearing, industrial, heavy-duty, nonlubricated parts. These materials are also sometimes referred to as "super build-up" alloys. If a machined finish is required, three layers are generally applied, with the assumption that the top layer will be removed during machining. A single layer is usually inadequate because of the effects of substrate/overlay intermixing (dilution).

Typical compositions and properties of the metal-to-metal wear alloys are presented in Tables 1 and 2. As with the build-up alloys, there appears to be no correlation between room-temperature hardness and abrasion properties.

In hostile environments, a high chromium content is beneficial. ER420 (American Welding Society, AWS, classification), and modified versions containing nickel, molybdenum, and niobium (or vanadium), are therefore the natural choice when high temperatures and mildly corrosive environments are encountered. For applications using steel mill hot-work rolls (which demand considerable hot hardness, resistance to oxidation, and resistance to thermal fatigue) both ER420 and EFe3 have been found suitable. Other applications for the metal-to-metal wear alloys in Table 2 include tractor rollers and crane wheels (EFe2), pincer guide shoes (EFe3), and blast furnace bells (ER420).

Metal-to-Earth Abrasion Alloys

The high-chromium irons encompass a wide range of compositions in which chromium may vary between approximately 6 and 35 wt%, and carbon may vary from about 2 to 6 wt%. Other possible alloying additions include molybdenum, manganese, and silicon.

The most important microstructural feature in the high-chromium irons, at least from a wear standpoint, is an M_7C_3 carbide, which forms in abundance during solidification and contains chromium, iron, and (if present) molybdenum. The matrix around these carbide particles can be austenitic, pearlitic, or martensitic (Ref 2). In general, the austenitic alloys rely on manganese for austenite stability.

As mentioned in the introduction of this article, carbon content is a good indicator of abrasion resistance for this class of materials. To illustrate this fact, low- and high-stress data are plotted for selected alloys in Fig. 1. This information was generated by the use of six different alloys (open arc/flux-cored wires) and nine different sets of welding parameters (Ref 3). The carbon contents referred to in the figures are those of the second layer of the two-layer overlays tested. The use of a shaded zone, rather than a line, indicates that there was considerable scatter in the data (particularly at the lower carbon contents, in the case of high-stress abrasion). The test methods and parameters were identical to those described in the section "Build-Up Alloys" in this article.

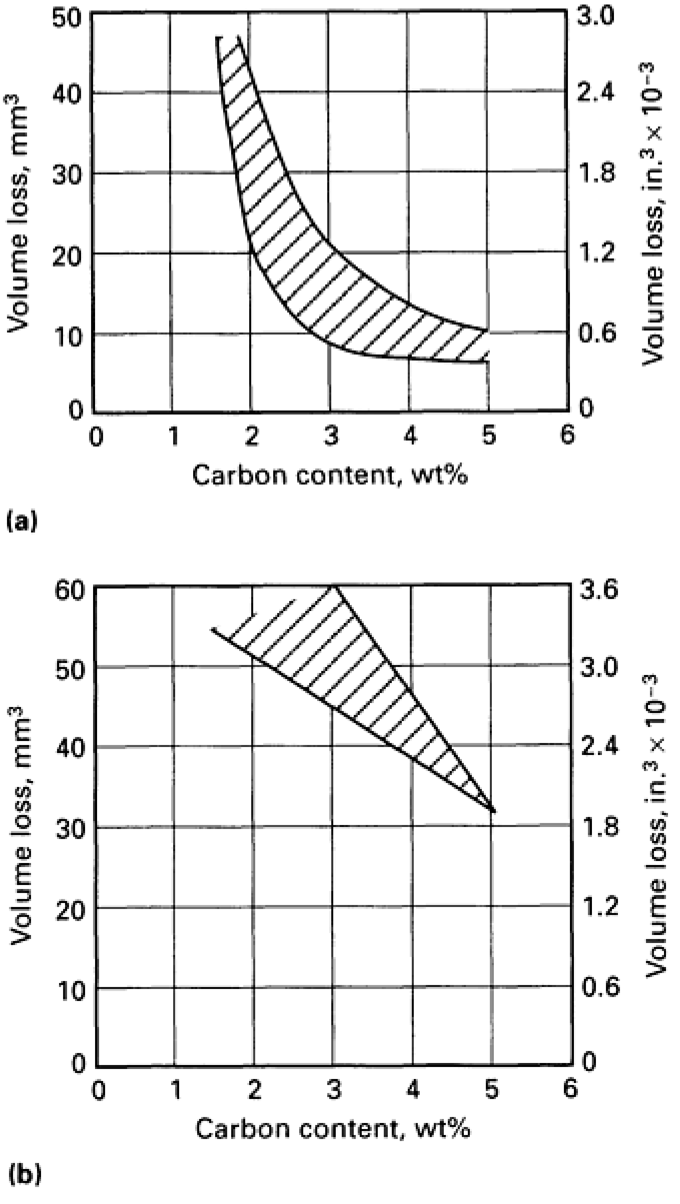


Fig. 1 Plot of volume loss versus carbon content for high-chromium iron metal-to-earth abrasion alloys. (a) Low-stress condition. (b) High-stress condition. Source: Ref 3

At high carbon and chromium levels, the formation of a hypereutectic microstructure, containing large, spinelike carbide particles (with a hexagonal cross section), is favored. At lower carbon and chromium contents, the microstructure is hypoeutectic. The microstructures of three-high-chromium irons (as deposited by the open arc welding process) are shown in Fig. 2. The nominal compositions of the same three alloys are presented in Table 3. As can be seen, ERFeCr-A3 (at a chromium content of 11 wt% and a carbon content of 2.6 wt%) exhibits a hypoeutectic (primary austenite) microstructure. The other two alloys, ERFeCr-A4(Mod) (29Cr/3.5C) and ERFeCr-A2 (28Cr/4.3C), possess a hypereutectic (primary carbide) structure. In Fig. 2, the large, spinelike carbides are shown in cross section.

Table 3 Composition of metal-to-earth abrasion alloys

Alloy	Composition, wt%							
	Cr	**C**	**Si**	**Mn**	**Mo**	**Ni**	**B**	**Fe**
ERFeCr-A3	11	2.6	1.3	1.8	1.5	. . .	. . .	bal
ERFeCr-A4(Mod)	29	3.5	1.1	0.9	. . .	2.6	0.7	bal

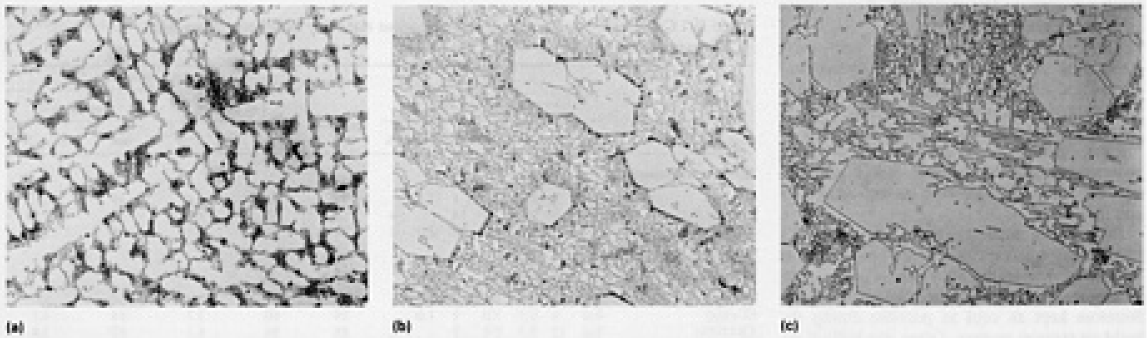


Fig. 2 Microstructure of high-chromium iron metal-to-earth abrasion alloys hardfaced with two-layer open arc deposit welding process. (a) ERFeCr-A3. (b) ERFeCr-A4(Mod). (c) ERFeCr-A2. 300×

In addition to M_7C_3, deposits of ERFeCr-A2 contain small quantities of M_6C and deposits of ERFeCr-A4(Mod) contain small quantities of both M_6C and M_3C. With regard to the matrix, both ERFeCr-A3 and ERFeCr-A4(Mod) exhibit a face-centered cubic (fcc) austenitic structure as deposited. ERFeCr-A2 is largely austenitic but may also contain small quantities of ferrite or martensite. The presence of martensite in the matrix structure of a high-chromium iron is believed to be beneficial to its high-stress abrasion resistance because of the additional support it provides to surface carbides.

For a given hardfacing consumable, deposit (overlay) microstructure is strongly influenced by the welding process and parameters used. Factors of concern, in this regard, include:

- Overlay/substrate intermixing
- Changes in composition due to losses in the arc
- Deposit cooling rate

With regard to industrial applications of the metal-to-metal abrasion alloys, the low-carbon (2 to 3% C) hypoeutectic materials are usually selected for situations involving moderate abrasion and impact, whereas the higher carbon (4 to 6% C) hypereutectic alloys are used in applications involving severe abrasion and little or no impact. Specific uses include:

- Crushers cones and pump casings (ERFeCr-A3)
- Bulldozer blades and crusher hammers [ERFeCr-A4(Mod)]
- Coal pulverizer liners and gravel pumps (ERFeCr-A2)

Tungsten Carbides

In contrast to the other weld overlay materials, the tungsten carbide composites do not rely upon the formation of suitable hard phases during weld pool solidification. Instead, these overlay materials rely on the transfer of tungsten carbide particles from the welding consumable to the overlay. It is important, therefore, to limit the heat input of the welding process in order to prevent melting of the tungsten carbide particles. If the tungsten carbide particles melt, they mix with

iron to form much softer iron-tungsten carbides, thus reducing abrasion resistance. For this reason, oxyacetylene deposits usually exhibit higher abrasion resistance than arc-welded tungsten carbide overlays.

An advantage of the tungsten carbide composites is that the size of the hard particles in the overlay can be controlled. This is important because abrasion resistance is dependent upon the size relationship between microstructural features (such as carbides) and the abrading particles. It is believed that, if the abrading particles are large in comparison to the microstructural particles, then, after a running-in period (during which the softer matrix material at the surface is worn down), the abrading particles ride over the hard microstructural outcrops. On the other hand, if the abrading particles are small in comparison to the microstructural particles, it is believed that the opportunity exists for wear of the matrix around the microstructural particles. Eventually, these may drop out, having played only a small in resisting abrasion.

The tungsten carbide particles themselves are typically manufactured by the melting of tungsten and carbon in a graphite crucible. During subsequent cooling, a two-phase mixture of WC and W_2C is formed. After crushing and screening (size selection), this material is incorporated into a tabular welding product with a carbon steel sheath.

Several tungsten carbide composites are available in a variety of tabular product forms. Popular compositions are 38, 50, 55, and 60 wt% tungsten carbide, with the carbon steel tube making up the balance. For each composition, several carbide size ranges are available. As an example, for the 60% WC oxyacetylene welding consumable, four mesh size ranges are available:

AWS designation	Mesh size range
RWC-12/20	12-20
RWC-20/30	20-30
RWC-30/40	30-40
RWC-40/120	40-120

The same composition is also available in flux-coated form for shielded metal are welding (SMAW) and as a continuous wire (with an internal flux) for open arc welding.

Tungsten carbide composites generally possess very high resistance to abrasion and very low impact strength. Performance in a given situation is dependent upon:

- Carbide volume fraction
- Size relationship between the carbides and the abrasive medium
- Welding technique applied

Important factors are the distribution of carbides in the overlay (because the particles tend to sink, turbulence in the molten weld pool is an advantage), and the amount of carbide dissolution and reprecipitation in the steel matrix during welding. Impact strength generally decreases with increasing carbide volume fraction.

Table 4 relates tungsten carbide composites and the previously mentioned alloys on the basis of low-stress and high-stress abrasion data (generated under identical test conditions). These values should be compared with the values in Table 2 and Fig. 1. Photomicrographs of the test deposits are shown in Fig. 3. Noteworthy features include:

- The difference in size of the tungsten carbide particles in the two materials
- The extent to which secondary carbides have precipitated within the matrix

Table 4 Abrasion data for tungsten carbide composites

Material		Abrasion, volume loss			
		Low-stress[(a)]		**High-stress**[(b)]	
Carbide, wt%	**Mesh size**	**mm^3**	**in.^{3} × 10^{-3}**	**mm^3**	**in.^{3} × 10^{-3}**
60	20-30	7.3	0.45	28.7	1.75
61	100-250	10.6	0.65	24.4	1.49

(a) Dry sand/rubber wheel test (ASTM G 65, Procedure B): load 13.6 kg (30 lb); 2000 rev.

(b) Slurry/steel wheel test (ASTM B 611, modified): load 22.7 kg (50 lb); 250 rev

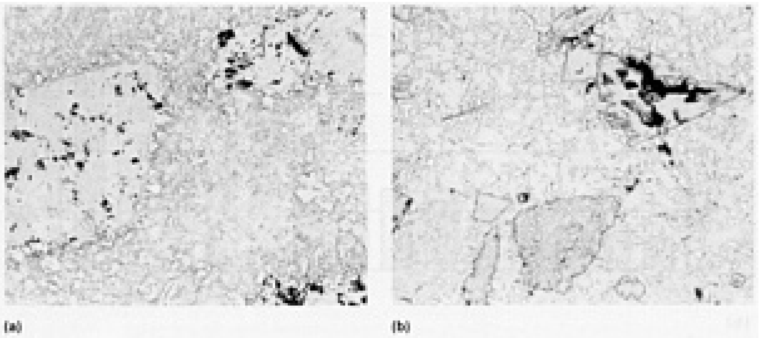


Fig. 3 Microstructures of tungsten carbide composites with carbides of different size in the weld overlay material. (a) 60% WC, 20 to 30 mesh particles. (b) 61% WC, 100 to 250 mesh particles. Overlay process applied is one-layer shielded metal arc (SMA) deposit. 120×

The tungsten carbide composites have been used to solve a wide variety of industrial sliding and drilling abrasion problems. Some of the common applications are plowshares, ditchdigger teeth, ripper teeth, and oil well drilling tools. For extremely hostile environments, some nonferrous tungsten carbide products (cobalt- and nickel-base products in the form of bare cast rods) are available. Also, several alternative composite materials, utilizing other carbides (for example, vanadium, titanium, or niobium), are available that have the advantage of creating a more homogeneous deposit because of their lower densities.

Nonferrous Alloys

Nonferrous hardfacing alloys are used either for high resistance to specific types of wear (other than abrasion), or for wear resistance (including abrasion) in environments that are too corrosive or beyond the service temperatures of ferrous alloys. The cobalt-base alloys and bronzes are particularly resistant to galling and to those wear processes involving microfatigue as the degradation mechanism (such as cavitation erosion). The cobalt-base alloys possess high resistance to deformation at temperatures in excess of 750 °C (1380 °F).

Cobalt-Base/Carbide-Type Alloys. Table 5 lists typical compositions of the nonferrous hardfacing alloys. The chief difference between the various cobalt-base/carbide-type alloys is in carbon content (hence, carbide volume fraction, room-temperature hardness, and level of abrasion resistance). Chromium-rich M_7C_3 is the predominant carbide in these alloys, although tungsten-rich M_6C is evident in those alloys having a high tungsten content, and chromium-rich $M_{23}C_6$ is common in the low-carbon alloys.

Table 5 Composition of selected nonferrous alloys

Alloy	Composition, wt%										
	Fe	**Cr**	**Mo**	**W**	**Si**	**C**	**B**	**Al**	**Cu**	**Co**	**Ni**
Cobalt-base/carbide type											
ERCoCr-A	. . .	28	. . .	5	. . .	1.2	. . .	. . .	. . .	bal	. . .
ERCoCr-B	. . .	29	. . .	8	. . .	1.5	. . .	. . .	. . .	bal	. . .
ERCoCr-C	. . .	31	. . .	13	. . .	2.5	. . .	. . .	. . .	bal	. . .
ERCoCr-E	. . .	27	6	. . .	. . .	0.2	. . .	. . .	. . .	bal	. . .
Cobalt- and nickel-base/Laves type											
T-400	. . .	9	29	. . .	2.5	. . .	. . .	. . .	. . .	bal	. . .
T-700	. . .	16	33	. . .	3.5	. . .	. . .	. . .	. . .	. . .	bal
T-800	. . .	18	29	. . .	3.5	. . .	. . .	. . .	. . .	bal	. . .
Nickel-base/boride type											
Alloy 40	1.5	7.5	. . .	. . .	3.5	0.3	1.5	. . .	. . .	. . .	bal
ERNiCr-B	3	11	. . .	. . .	4	0.5	2.5	. . .	. . .	. . .	bal
ERNiCr-C	4	16	. . .	. . .	4	0.7	3.5	. . .	. . .	. . .	bal
Aluminum bronze type											
ECuAl-B	4	. . .	. . .	. . .	1.0	. . .	. . .	9	bal	. . .	. . .
ECuAl-D	4	. . .	. . .	. . .	. . .	. . .	. . .	13.5	bal	. . .	. . .

Low-stress abrasion, galling, and cavitation erosion data are presented for many of the nonferrous hardfacing alloys in Fig. 4, 5, and 6. Figure 4 relates the carbon content and the resistance to low-stress abrasion for the cobalt-base/carbide-type alloys (ERCoCr-A, -B, -C, and -E). From Fig. 5 and 6, it is evident that the cobalt-base matrix (solid solution) is responsible for the excellent self-mated sliding properties and cavitation erosion resistance of these alloys. Additional information is available in the article "Friction and Wear of Cobalt-Base Wrought Alloys" in this Volume.

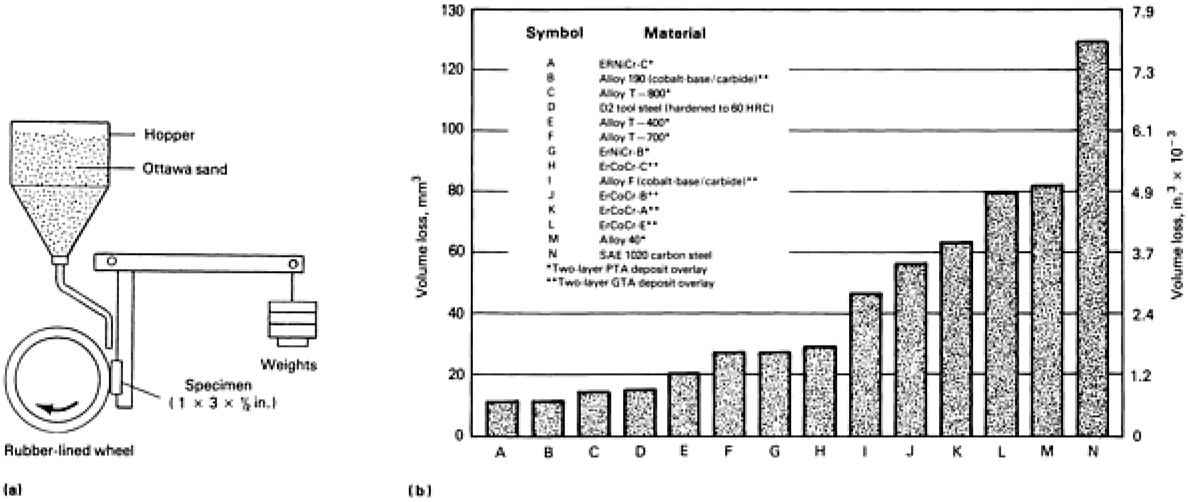


Fig. 4 Comparison of hardfaced nonferrous alloys to tool steel and carbon steel reference materials using ASTM G 65 low-stress abrasion test. (a) Schematic of G 65 dry sand/rubber wheel test apparatus. (b) Low-stress abrasion test data. G 65 test parameters: procedure B; room temperature; 13.6 kg (30 lb) load; quartz grain sand diameter of 212 to 300 μm (8.48 to 12 mil); 2000 rev at 200 rev/min; 390 g/min (0.86 lb/min) feed rate

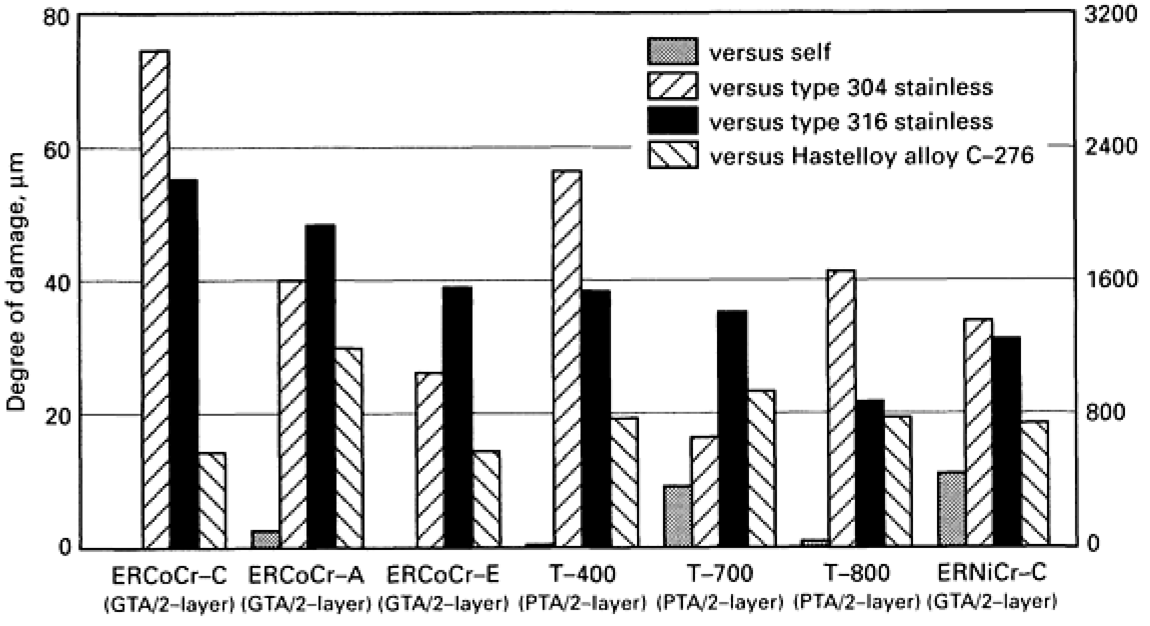


Fig. 5 Galling property evaluation of nonferrous alloys versus selected alloys (itself, stainless steels, and a nickel-molybdenum-chromium alloy). Data obtained using pin-on-block test with following parameters: 2722 (6000 lb) load; 10 strokes through 120° arc. GTA, gas tungsten arc

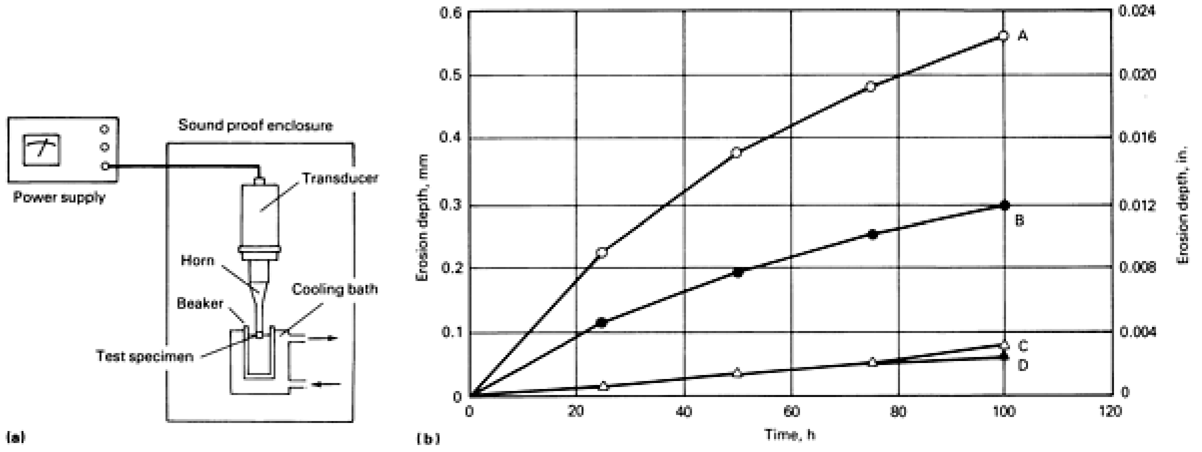


Fig. 6 Comparison of hardfaced nonferrous alloys to a stainless steel and a nickel-molybdenum chromium alloy using ASTM G 32 vibratory cavitation test. (a) Schematic showing G 32 test apparatus, (b) Cavitation erosion test data. G 32 test parameters: medium, distilled water; test temperature, 16 °C (61 °F); vibration, 20 kHz frequency at 0.05 mm (0.002 in.) amplitude. A, type 410 stainless steel (hardness, 23 HRC); B, Hastelloy alloy C-276; C, ERCoCr-E (two-layer GTA deposit); D, ERCoCr-A (two-layer GTA deposit)

The low-stress abrasion data in Fig. 4 were generated using the ASTM G 65 dry sand/rubber wheel test. The vibratory cavitation test described in ASTM G 32 was used to generate the results given in Fig. 6. The results in Fig. 5 relate to a pin-on-block galling test (Ref 4).

Photomicrographs of plasma-transferred arc (two-layer) deposits of three cobalt-base/carbide-type alloys are shown in Fig. 7. The differences in carbide volume fraction and geometry are evident from these photomicrographs.

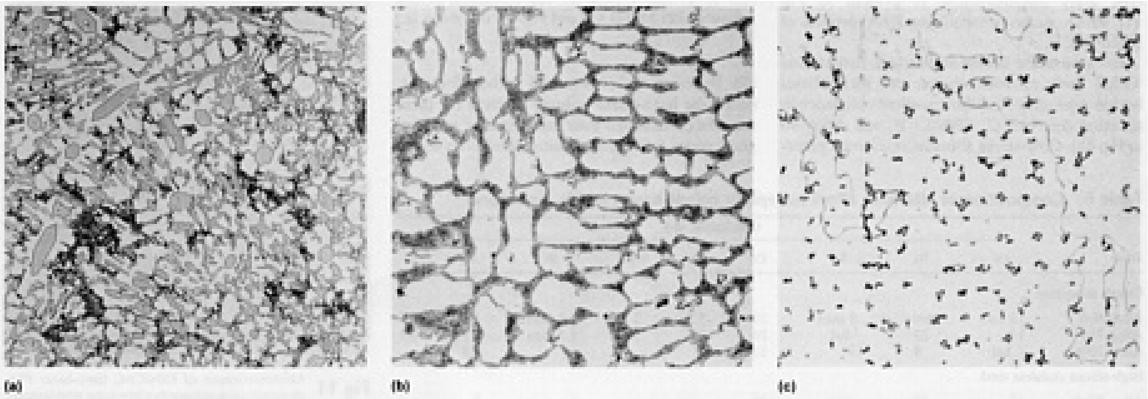


Fig. 7 Microstructures of two-layer PTA deposited hardfacing on cobalt-base/carbide type alloys. (a) ERCoCr-C. (b) ERCoCr-A. (c) ERCoCr-E. 425×

Iron and Nickel Substitutes for Cobalt Compositions. Because of the relatively high cost of cobalt, attempts have been made to design alternate materials, with iron and nickel as the predominant cobalt substitutes. The compositions of three such materials are given in Table 6. A comparison of the sliding wear properties of the cobalt-base/carbide-type alloys and these three alternate alloys, as a function of temperature up to 1000 °C (1830 °F), is presented in Fig. 8. From the data in Fig. 8, it is evident that cobalt content is a very important factor up to ~750 °C (~ 1380 °F), above which oxide glaze formation is observed.

Table 6 Composition of alternate alloys to replace cobalt-base/carbide-type alloys

Alloy	Composition, wt%								
	Co	**Ni**	**Fe**	**Cr**	**Mo**	**W**	**Si**	**C**	**B**
Cobalt substitute									
Alloy N-6	3 max	bal	3 max	29	5.5	2	1.5 max	1.1	0.6
Alloy 716	11	23	bal	26	3	3.5	1.5 max	1.1	0.5
Alloy 2006	bal	8	18	31	8	. . .	1	1.3	. . .
High-silicon stainless steel									

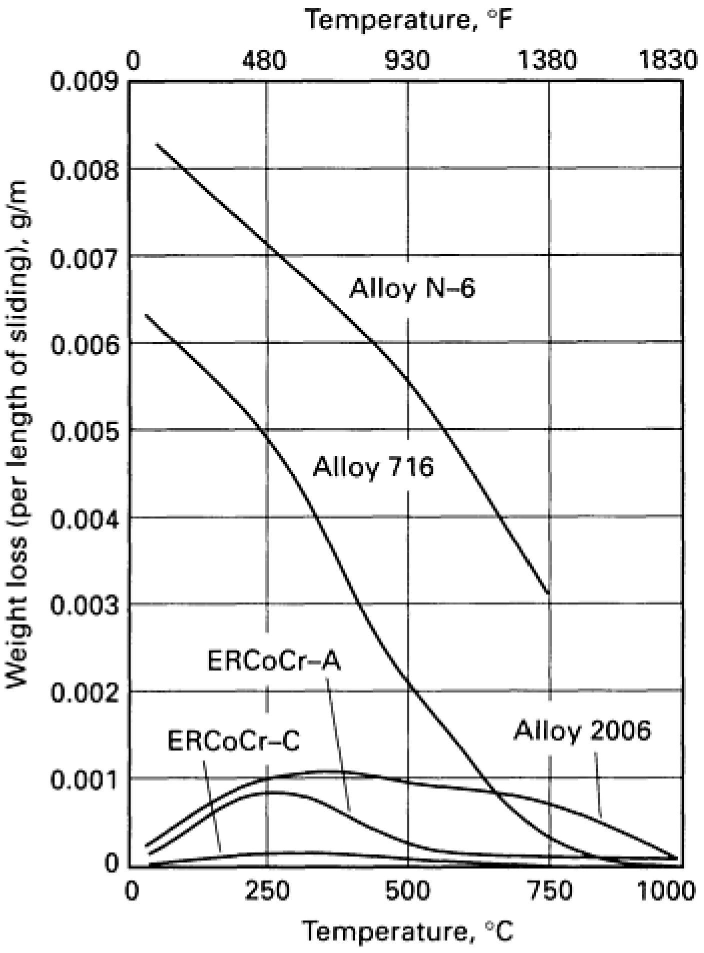


Fig. 8 Plot of wear rate versus temperature to compare sliding properties of cobalt-base/carbide-type alloys and selected alternate alloys, all of which have been GTA deposit hardfaced. Test conditions include pressure of 21 MPa (3 ksi) and velocity of 8 mm/s (0.3 in./s).

The information in Fig. 8 is also useful because it indicates a maximum sliding wear rate for the cobalt-base hardfacing alloys (self-mated) at ~250 °C (~480 °F). During these (reciprocating) sliding wear tests (Ref 5), the coefficient of friction was monitored continuously. Not only was this parameter found to vary through each sliding cycle, but the values also drifted with test time (hence sliding distance). For reference, friction coefficient traces for ERCoCr-A, at 25, 500, and 1000 °C (75, 930, and 1830 °F), are shown in Fig. 9. Note that relatively high values were obtained in the early stages of testing at 1000 °C (1830 °F) (presumably during the time interval in which the oxide glaze was forming).

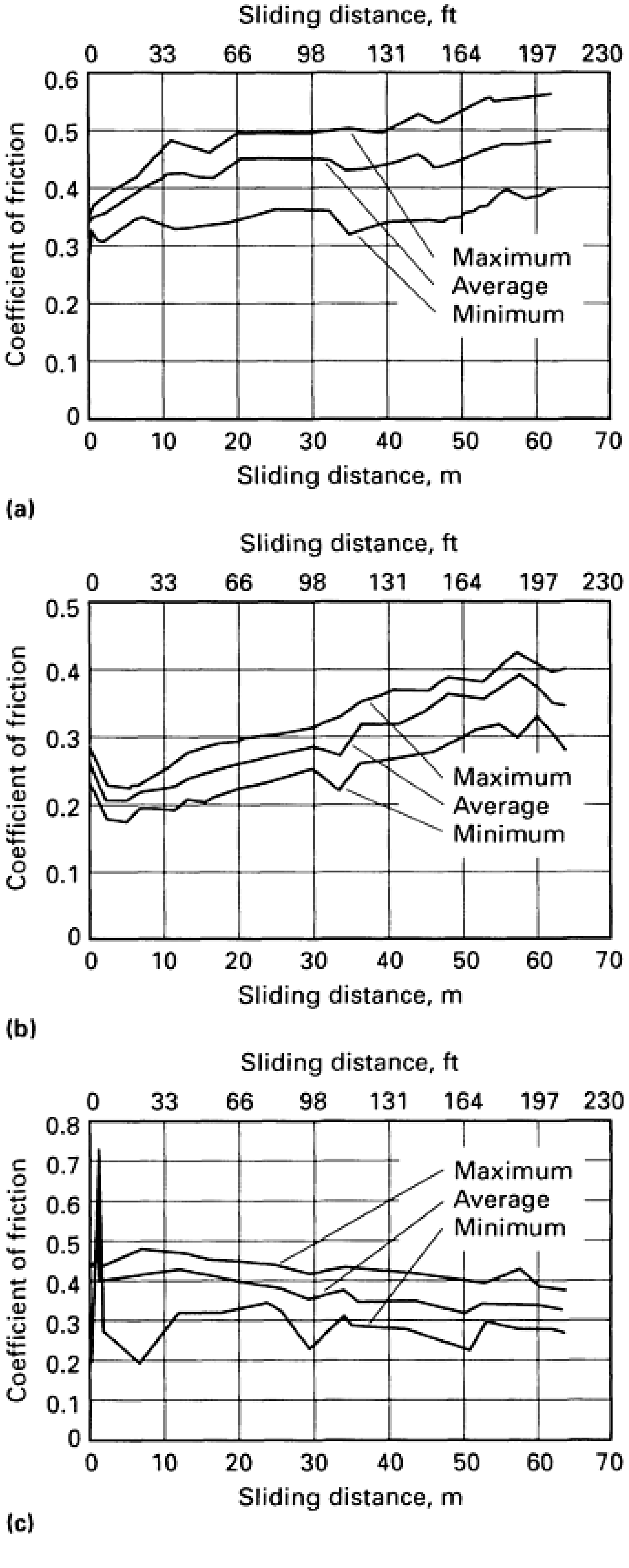


Fig. 9 Plot of coefficient of friction versus sliding distance for self-coupled ERCoCr-A (GTA deposit) hardfaced cobalt-base/carbide-type alloy as a function of test temperature. (a) 25 °C (75 °F). (b) 500 °C (930 °F). (c) 1000 °C (1830 °F). Test conditions include pressure of 21 MPa (3 ksi) and velocity of 8 mm/s (0.3 in./s).

High-Silicon Stainless Steel Alternate Material. Recently, several new high-silicon stainless steels have been introduced that may be considered as alternate materials. These exhibit identical anti-galling characteristics and possess equal resistance to cavitation erosion. Some of these have been developed for general use, and others for specific applications (for example, hydroelectric turbines and nuclear valves). Some are based on type 200 (manganese- and nitrogen-containing) austenitic stainless steel; others are based on type 300 stainless steel. One of these latter compositions is presented in Table 6. These high-silicon hardfacing alloys differ from the traditional cobalt-base/carbide-type alloys in terms of their corrosion resistance (they are better in some aqueous media and worse in others), their mechanical properties (particularly at high temperatures), and their thermal stability.

Laves-Type Alloy Compositions. Three cobalt- and nickel-base/Laves-type alloy compositions are presented in Table 5. In these materials, molybdenum and silicon are added at levels in excess of their solubility limit with the objective of inducing the precipitation of the hard (and corrosion-resistant) Laves phase (an intermetallic compound). Carbon is held as low as possible in these alloys to discourage carbide formation. The microstructure of alloy T-800 (two-layer PTA deposit) is shown in Fig. 10.

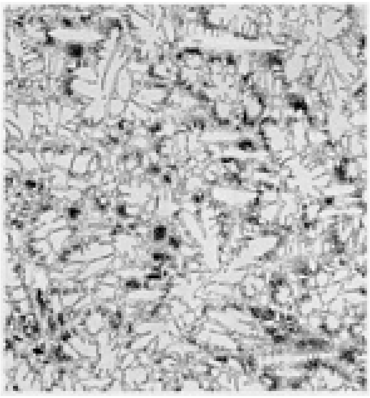


Fig. 10 Microstructure of Alloy T-800 (two-layer PTA deposit) cobalt-base/Laves-type nonferrous alloy. 200×

Because the Laves intermetallic phase is so abundant in these alloys, its presence governs all the material properties. Accordingly, the effects of the matrix composition in these alloys are less pronounced than is the case for the cobalt-base/carbide-type alloys, for example. The Laves phase is especially responsible outstanding abrasion resistance, but it severely limits the material ductility and the impact strength. In fact, it is difficult to attain crack-free overlays on all but the smallest components given adequate preheat. For this reason, these alloys have been more successful as thermal spray materials.

Nickel-Base/Boride-Type Alloys. Of all the hardfacing alloys, the nickel-base/boride-type alloys are microstructurally the most complex. The alloy compositions represent a progression in terms of iron, chromium, boron and carbon contents (Table 5). Iron content is largely incidental, allowing the use of ferrocompounds during manufacture. Together with nickel, the other three elements determine the level and type of hard face within the structure upon solidification, boron being the primary hard-face forming element (for which nickel and chromium compete) and carbon being the secondary hardphase former).

The chief purpose of silicon in the material is to provide, in conjunction with boron, self-fluxing characteristics. However, as an important matrix element and as a potential promoter of intermetallic precipitates, it also has a powerful influence on the wear properties of the alloys.

The actual phases that form in the nickel-base/boride-type alloys are listed in Table 7 on the basis of chromium content.

Table 7 Phases formed in nickel-base/boride-type alloys

Chromium content	Secondary phases		Dominant hard phase
	Compound formed	**Required conditions**	
Low (~5 wt%)	Ni_3Si	>3 wt% Si	Ni_3B
Medium (~15 wt%)	Ni_3Si	>2.5 wt% Si	Ni_3B and chromium boride (usually CrB, although Cr_2B and Cr_3B_2 may also be present)
High (~25 wt%)	Ni_3Si	>3 wt% Si	CrB and Cr_5B_3
All levels	. . .	. . .	Complex carbines of $M_{23}C_6$ and M_7C_3 types

Boron content influences the level of silicon required for silicide (Ni_3Si) formation. The higher the boron content, the lower is the silicon content required to form silicides.

The complexity of the nickel-base/boride-type alloy microstructure is evident in Fig. 11, which shows a two-layer PTA deposit of ERNiCr-C.

Fig. 11 Microstructure of ERNiCr-C (two-layer PTA deposit) nickel-base/boride-type nonferrous alloy. 425×

Because of the boride and carbide dispersions within their microstructures, the nickel-base/boride-type alloy exhibit excellent resistance to abrasion (ERNiCr-C, ERNiCr-B, and Alloy 40 in Fig. 4b). Low-stress abrasion resistance generally increases with boron and carbon contents, hence the hard-phase volume fraction for these materials. Although their performance is not as good under self-mated sliding as the cobalt-base materials, the nickel-base/boride-type alloys possess moderate resistance to galling (Fig. 5). Of the nonferrous materials, the nickel-base/boride-type alloys are the least resistant to corrosion. This is attributed to the lack of chromium in the matrix that follows boride and carbide formation.

Originally, the nickel-base/boride-type alloys were developed from brazing alloy compositions specifically for use with the spray-and-fuse process (for which their self-fluxing properties are ideal). During the fusing cycle of this process, it is believed that oxides within the sprayed coating combine with some of the silicon and boron to form a borosilicate slag (which floats to the surface of the deposit). The only other commercially significant spray-and-fuse alloys are some silicon-containing and boron-containing derivatives of the cobalt-base/carbide-type materials. Microstructurally, and in a property sense, these alloys bear little resemblance to their traditional counterparts (because the presence of boron markedly changes the matrix chromium content and the types of hard phase that form during solidification.

Bronze Type. The two bronze compositions given in Table 5 fall in the aluminum bronze category. Silicon and other types of bronze are also available in welding consumable form for hardfacing. Aluminum, silicon, and iron are added to the bronzes to strengthen the solid solution, and, in excess of their solubility limits (about 8%, in the case of aluminum), to provide precipitation hardening.

The high resistance to sliding wear and cavitation erosion of these alloys is attributed to planar slip during deformation. This is believed to delay the onset and progression of microfatigue (as in the case of the cobalt-base alloys). Wear tests have shown that the aluminum bronzes possess anti-galling characteristics (self-coupled) comparable to those of ERCoCr-E; low-stress abrasion resistance of the aluminum bronzes, however, was very low.

Components typically protected using the aluminum bronzes include gears, cams, and cold drawing/forming dies. These alloys are not recommended for elevated-temperature use because their mechanical properties considerably decrease at temperatures >200 °C (>390 °F).

References

1. D.K. Subramanyam, A.E. Swansiger, and H.S. Avery, Austenitic Manganese Steels, *Properties and Selection: Irons, Steels, and High-Performance Alloys,* Vol 1, *Metals Handbook,* 10th ed., 1990, p 822
2. H.S. Avery and H.J. Chapin, *Weld. J.,* Vol 31, 1952, p 917
3. R.D. Zordan, Report 10890, Cabot Corporation, 1982
4. P. Crook and T.H. Shailes, *Surfacing J.,* Vol 16, 1985, p 35
5. P. Crook and C.C. Li, *Wear of Materials,* American Society of Mechanical Engineers, 1983, p 272

Friction and Wear of Cobalt-Base Wrought Alloys

Paul Crook, Haynes International, Inc.; Alan V. Levy, Lawrence Berkeley Laboratory

Introduction

COBALT-BASE WROUGHT ALLOYS can conveniently be categorized as follows:

- Alloys with high-carbon contents and designed for wear service
- Alloys designed for high-temperature service
- Alloys designed to combat simultaneous corrosion (aqueous) and wear

The high-carbon alloys, such as Alloy 6B and Alloy 6K, are essentially wrought versions of the weld overlay materials described in the article "Friction and Wear of Hardfacing Alloys" in this Volume. Wrought processing improves chemical homogeneity (important in a corrosion sense), markedly increases ductility, and modifies substantially the geometry of the carbide precipitates within the alloys (which in turn enhances resistance to low-stress abrasion). In terms of composition (Table 1), the alloys are essentially Co-Cr-W-C quaternaries with chromium providing strength and corrosion resistance to the solid solution in addition to functioning as the chief carbide former (during ally solidification). Tungsten provides additional solid solution strength. The attractive wear properties of these alloys are attributed chiefly to the unusual deformation characteristics imparted by cobalt. The carbides play a secondary role, being particularly important under conditions conducive to low-stress abrasion. With regard to the influence of cobalt, the low stacking fault energy of the (high-temperature stable) face-centered cubic (fcc) structure is important, as is the transformation to the (low-temperature stable) hexagonal close-packed (hcp) form under the action of mechanical stress (Ref 1, 2, 3, 4). The former is believed to delay fatigue crack initiation (a factor in many wear processes) through its influence on slip behavior and twinning propensity; the latter is believed to enhance fatigue life by reducing internal strain.

Table 1 Nominal compositions of wrought cobalt-base alloys

Alloy	Composition, wt%									
	Cr	**W**	**Mo**	**C**	**Fe**	**Ni**	**Si**	**Mn**	**Others**	**Co**
Wear-resistant alloys										
Alloy 6B	30	4	1.5 max	1	3 max	2.5	0.7	1.4	. . .	bal
Alloy 6K	30	4.5	1.5 max	1.6	3 max	3 max	2 max	2 max	. . .	bal
High-temperature alloys										
Alloy 25	20	15	. . .	0.1	3 max	10	0.4 max	1.5	. . .	bal
Alloy 188	22	14	. . .	0.1	3 max	22	0.35	1.25	0.03 La	bal
Corrosion-resistant alloys										
Ultimet	26	2	5	0.06	3	9	0.3	0.8	0.08 N	bal
MP 159	19	. . .	7	. . .	9	25.5	. . .	. . .	3 Ti, 0.6 Nb, 0.2 Al	bal
MP35N	20	. . .	10	. . .	. . .	35	. . .	. . .	. . .	bal

The high-temperature alloys, such as Alloy 25 (L-605) and Alloy 188, are considerably more ductile, oxidation resistant, and microstructurally stable. Both examples contain approximately 0.1 wt% C (about one-tenth of that in Alloy 6B), which is sufficient to provide carbide strengthening, yet low enough to maintain ductility. Structural stability is enhanced in these alloys by nickel, which decreases the fcc/hcp transformation temperature in cobalt-base alloys (Ref 5). Intermetallic formation in Alloy 188 is minimized by control of the electron hole number (the intermetallic Laves phase is a common precipitate in the cobalt-base alloys during long-term exposure to elevated temperatures). Alloy 188 also contains a small amount of the rare-earth element lanthanum for increased resistance to oxidation (Table 1).

Materials in the third, and final, category differ from those already mentioned in that they utilize molybdenum in preference to tungsten. Whereas tungsten appears to be a more effective strengthener at a elevated temperatures, molybdenum appears to impart a greater degree of resistance to a variety of wet corrosive media. In addition, in some of these alloys--for example, MP35N alloy, MP159 alloy, and Ultimet alloy (1233)--carbon is held within the soluble range to improve resistance to heat-affected zone (HAZ) sensitization during welding.

Typical microstructures of the wrought cobalt-base alloys are shown in Fig. 1. All represent the alloys in the solution-annealed condition (their typical condition of use, except in the case of MP35N and MP159 alloys, which are controlled by cold reduction and aging and may be purchased at various strength levels).

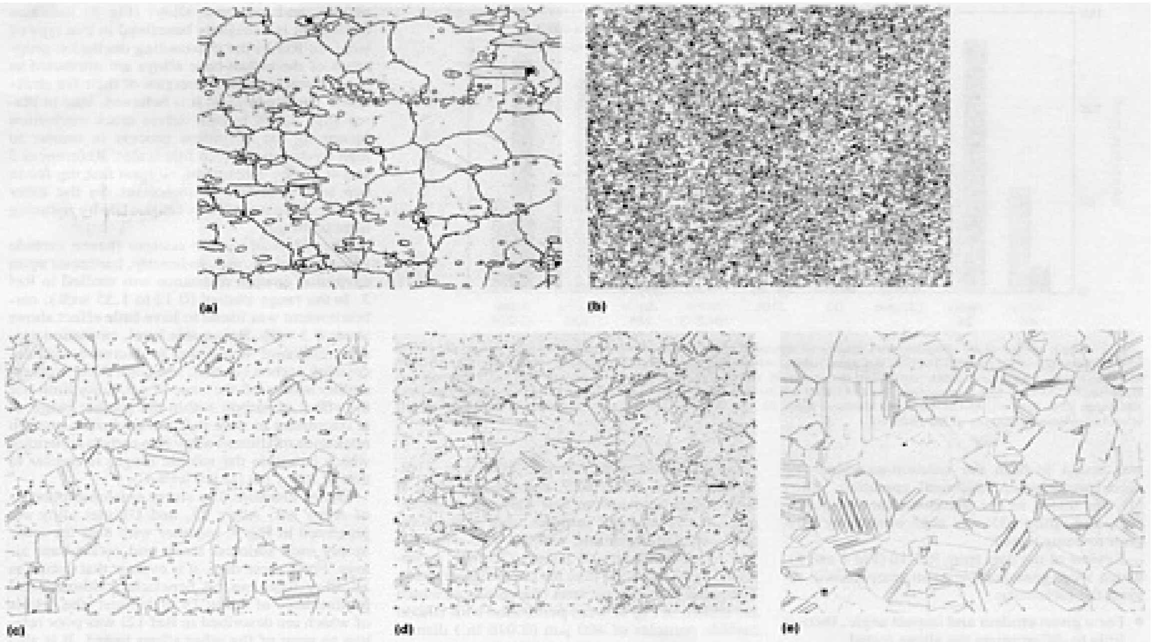


Fig. 1 Microstructure of wrought cobalt-base alloys etched with hydrochloric acid/oxalic acid solution. (a) Alloy 6B. (b) Alloy 6K. (c) Alloy 25. (d) Alloy 188. (e) Ultimet alloy. 200×

Alloy 6B (Fig. 1a) contains approximately 12.5 wt% carbides of the M_7C_3 and $M_{23}C_6$ types in the ratio 9:1 (Ref 3). Alloy 6K (Fig. 1b) exhibits an even greater carbide volume fraction, again with M_7C_3 as the predominant type (Ref 6). By virtue of much lower carbon contents, the high-temperature Alloy 25 (Fig. 1c) and Alloy 188 (Fig. 1d) contain appreciably less carbide precipitation. Nevertheless, this precipitation, which is predominantly of the M_6C type, is important to the high-temperature properties of these materials, partially because it restricts grain growth during heat treatment and service (Ref 7).

Ultimet alloy (Fig. 1e) is typical of the low-carbon corrosion-resistant alloys in that it exhibits no intragranular or grain boundary carbide precipitation, although a sparse dispersion of nitrides in evident in the structure.

In the solution-annealed condition, the wrought cobalt-base alloys exhibited a fcc matrix structure. With the exception of Alloy 188 (which is fcc stable), however, the alloys in Table 1 are metastable at room temperature and tend to transform to hexagonal close-packed under the action of mechanical stress or during heat treatment (at temperatures below the transformation temperature). Alloys 25, for example, is believed to be metastable at temperatures <760 °C (<1400 °F) (Ref 8).

Wear Properties

There are several distinct types of wear, and they generally fall into three main categories:

- Abrasive wear
- Erosive wear
- Sliding wear

Abrasive Wear

The low-stress and high-stress abrasive wear characteristics of three of the compositions in Table 1, relative to other common alloys, are shown in Fig. 2. Review of these (room temperature) data reveals:

- The importance of hard phases (carbides) within the microstructure, under conditions conducive to low-stress abrasion
- Their lack of importance under high-stress conditions (the performance of the carbide-containing Alloy 6B being similar to that of the carbide-free Ultimet alloy)

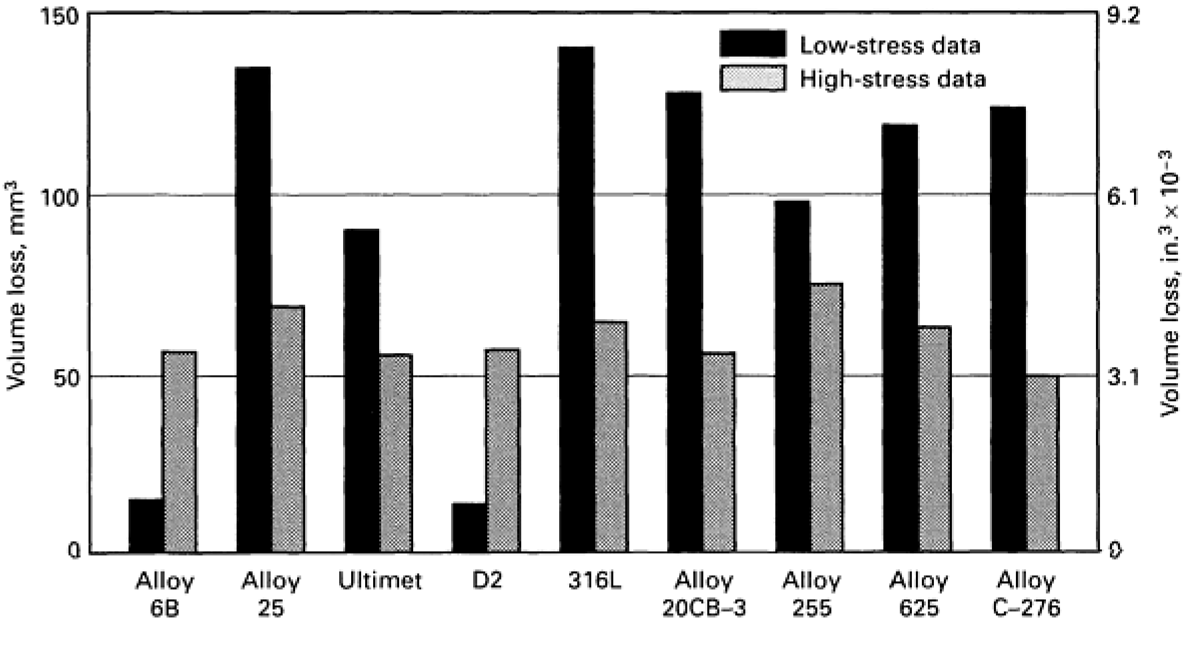


Fig. 2 Room-temperature abrasion test results of wrought cobalt-base alloys compared with other alloys. Solid

bars, low-stress data per ASTM G 65 dry sand/rubber wheel test (procedure B). Shaded bars, high-stress data per ASTM B 611 slurry/steel wheel test. Both tests used quartz sand of 212 to 300 μm (0.0085 to 0.012 in.) diameter as abrasive. ASTM G 65 test parameters: load, 13.6 kg (30 lb); wheel revolutions, 2000 rev at 200 rev/min; feed rate, 390 g/min (0.86 lb/min). ASTM B 611 parameters: load, 22.7 kg (50 lb); 1500 g (3.3 lb) sand to 940 g (2.1 lb) water; wheel revolutions, 250 rev at 245 rev/min

Low-Stress Abrasive Wear. The importance of the carbide precipitates, in particular their size and shape, under low-stress conditions can be appreciated by comparing the microstructures and volume losses for Alloy 6B and alloy 6 (the cast equivalent) in the article "Cobalt and Cobalt Alloys" in *Properties and Selection: Nonferrous Alloys and Special-Purpose Materials,* Volume 2 of *ASM Handbook,* formerly 10th Edition, *Metals Handbook.* Large blocky carbides have been found to be of most benefit under such conditions (Ref 6).

The low-stress abrasion results for Alloys 25, Ultimet alloy, the stainless steels, and the high-nickel alloys correlate well with the annealed hardness of the samples (Table 2) with the exception of Alloy 25.

Table 2 Annealed hardnesses of low-stress abrasion samples of wrought cobalt-base alloys relative to other alloys

Alloy	Hardness, HRC
Alloy 6B	38.1
Alloy 25	24.7
Ultimet	35.6
D2 tool steel	59.5
316L stainless steel	81.6[(a)]
Alloy 20CB-3	86.8[(a)]
Alloy 255	27.2
Alloy 625	20.2
Alloy C-276	89.8[(a)]

(a) HRB

High-Stress Abrasive Wear. For the high-stress test, no relationship between room-temperature hardness and performance is evident. In fact, all the alloys, including Alloy 6B and D2 tool steel, fall within a narrow performance band, suggesting that resistance to fracture is at least as important as resistance to deformation under such conditions. It is noteworthy that the duplex (austenitic-ferritic) microstructure of Alloy 255 is beneficial under low-stress conditions, but detrimental in the high-stress case.

Erosive Wear

The cobalt-base wrought alloys have been evaluated under conditions conducive to three distinct forms of erosive wear:

- Solid-particle impingement erosion
- Cavitation erosion
- Slurry erosion

Solid-particle Impingement Erosion. To define the behavior of the cobalt-base wrought alloys under solid-particle impingement erosion conditions, the results of two independent studies (using the ASTM G 76 test apparatus shown in Fig. 3) are presented in graphical form in Fig. 4, 5, 6, and 7. In the first study by Ninham (Ref 10), Alloy 6B and Alloy 188 were compared with Alloy C-276, Alloy 255, and 718 at three impact angles (30, 60, and 90°) and for two types of erodent (silicon carbide particles of 250 to 300 μm, or 0.010 to 0.012 in., diameter and angular quartz particles of 70 to 200 μm, or 0.003 to 0.008 in., diameter). To assess the influence of alloy condition, Alloy 188 and Alloy C-276 were tested in both the solution annealed and 30% cold-worked conditions, and Alloy 718 was tested in both the solution-annealed and aged (precipitation-hardened) condition. Alloy 6B was tested in the solution-annealed condition, and Alloy 255 was cold worked by 30% prior to being tested.

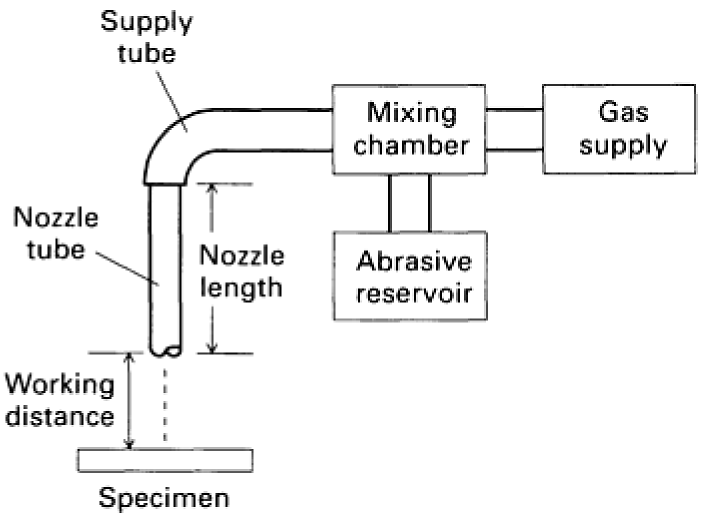


Fig. 3 Schematic showing key components of the ASTM G 76 solid-particle gas-jet impingement erosion test apparatus

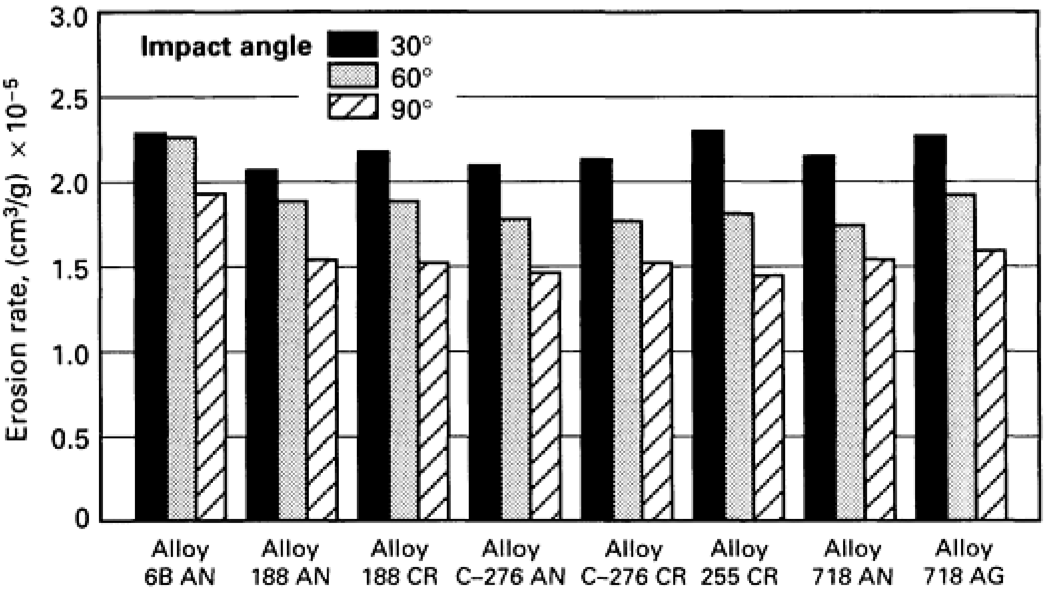


Fig. 4 Solid-particle impingement erosion test (ASTM G 76) results to compare cobalt-base alloys with selected alloys using a 250 to 300 μm (0.010 to 0.012 in.) diameter silicon carbide erodent at impact angles of 30, 60, and 90°. Tests conducted at room temperature with 60 m/s (200 ft/s) particle velocity. AN, annealed; CR, cold reduced; AG, aged. Source: Ref 10

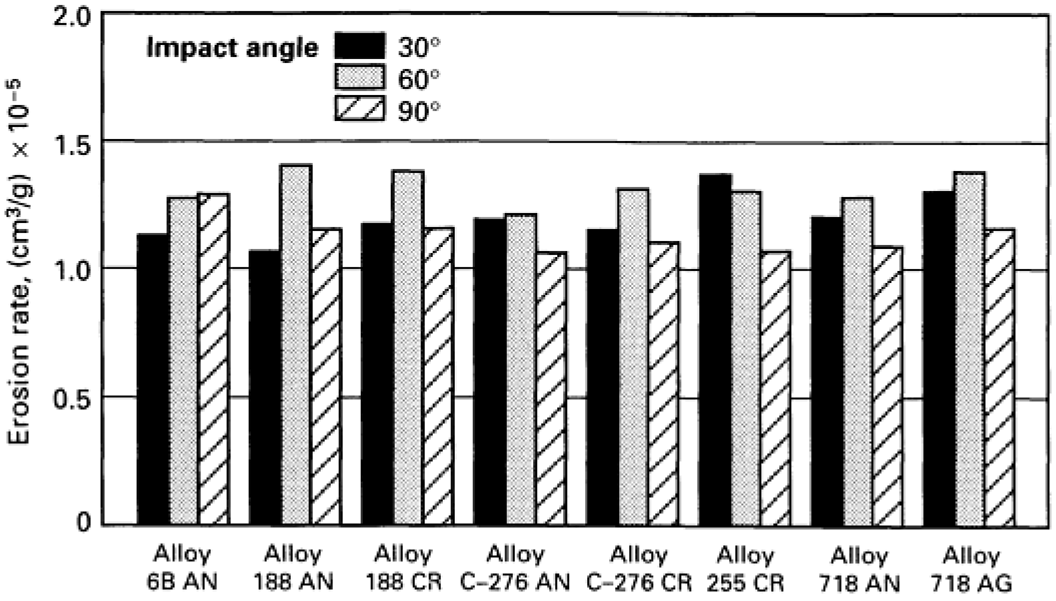


Fig. 5 Solid-particle impingement erosion test (ASTM G 76) results to compare cobalt-base alloys with selected alloys using a 75 to 200 μm (0.003 to 0.008 in.) diameter quartz erodent of impact angles of 30, 60, and 90°. Tests conducted at room temperature with 60 m/s (200 ft/s) particle velocity. AN, annealed; CR, cold reduced; AG, aged. Source: Ref 10

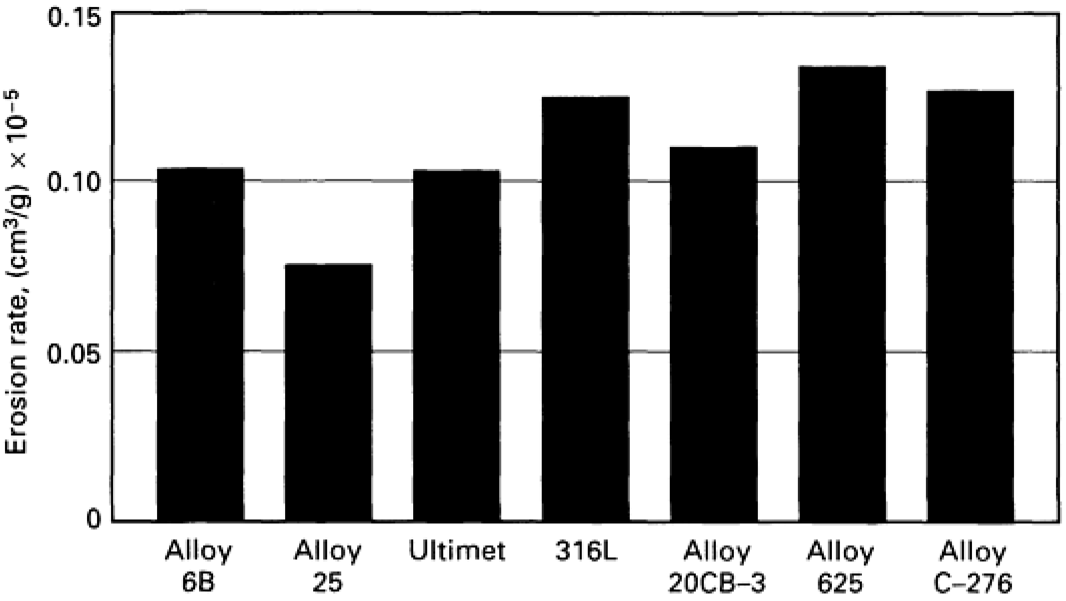


Fig. 6 Solid-particle impingement erosion test (ASTM G 76) results to compare cobalt-base alloys with selected alloys using a 400 μm (0.016 in.) mean diameter silicon carbide erodent at an impact angle of 60°. Tests parameters: test temperature, 20 °C (70 °F); particle velocity, 20 m/s (65 ft/s); particle feed rate, 0.02 g/s (4×10^{-5} lb/s)

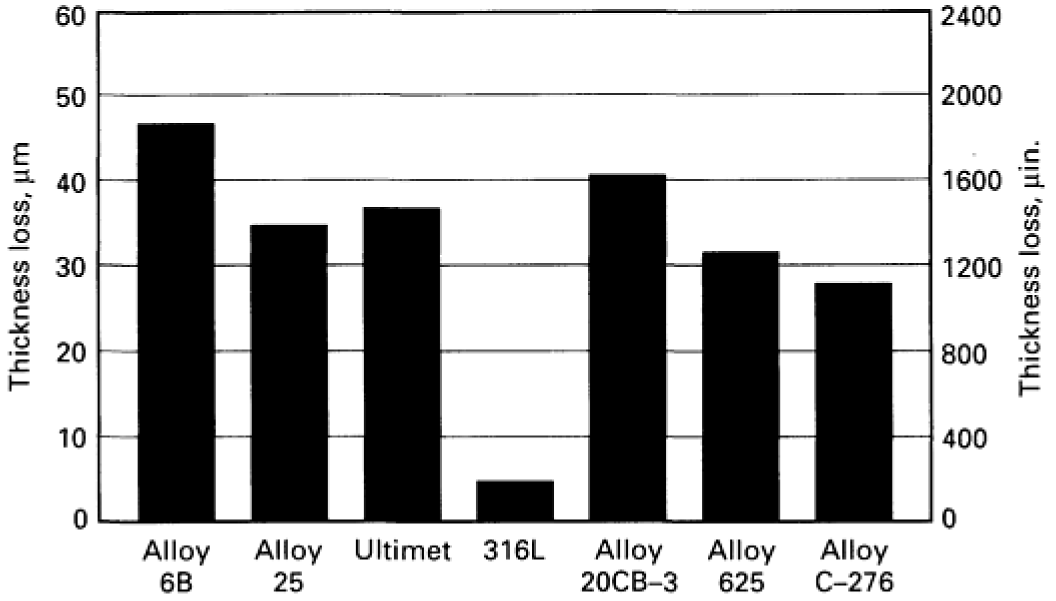


Fig. 7 Solid-particle impingement erosion test (ASTM G 76) results to compare cobalt-base alloys with selected alloys using an 80 μm (0.003 in.) mean diameter alumina erodent at an impact angle of 30°. Tests parameters: test temperature, 850 °C (1560 °F); particle velocity, 20 m/s (65 ft/s); particle feed rate, 0.02 g/s (4 × 10^{-5} lb/s)

Review of the data from Ref 10 (Fig. 4 and 5), which were generated at room temperature, reveals the following:

- For a given erodent and impact angle, there is little to differentiate the alloys tested
- Neither cold working nor precipitation hardening has a strong effect on solid-particle erosion resistance
- For silicon carbide, erosion rate decreases as impact angle increases, in the range 30 to 90°; this is consistent with other erosion studies of ductile materials (Ref 11)
- For quartz, this relationship does not apply because the materials tested have exhibited a maximum erosion rate at an impact angle of 60°; Alloy 6B is one of the two exceptions in this regard, having exhibited a maximum erosion rate at 90°

In the second solid-particle erosion study involving the cobalt-base wrought alloys by the authors, Alloy 6B, Alloy 25, and Ultimet alloy were compared with Alloy C-276, type 316L stainless, Alloy 625, and 20CB-3 alloy at both room temperature and elevated temperature. In this case, all samples were in the solution-annealed condition, and the tests were run at a particle velocity of 20 m/s (70 ft/s), as opposed to 60 m/s (200 ft/s) for the previous work. The room-temperature tests (the results of which are shown in Fig. 6) were performed with silicon carbide particles of 400 μm (0.016 in.) diameter, at an impact angle of 60°. The elevated-temperature (850 °C, or 1560 °F); tests were run with 80 μm (0.003 in.) diameter alumina particles at an impact angle of 30°; the results are shown in Fig. 7.

Under the conditions of the room-temperature test, the three cobalt-base alloys exhibited the lowest erosion rates. Alloy 25, in particular, was significantly more resistant to erosion than the other test alloys. At elevated temperature, however, 316L was by far the most erosion-resistant material, presumably due to the characteristics of the oxide scale during impingement.

Cavitation Erosion. Comparison of the cavitation erosion resistance of several cobalt-, nickel-, and iron-base alloys (Fig. 8) indicates that cobalt is extremely beneficial in this type of wear. In Ref 1, the outstanding cavitation properties of the cobalt-base alloys are attributed to the low stacking fault energies of their fcc structures. These energies, it is believed, lead to planar slip, which in turn delays crack nucleation (assuming the cavitation process is similar to high-cycle fatigue on a fine scale). References 3 and 4, on the other hand, suggest that the fcc to hcp transformation is important (in the latter case because it enhances fatigue life by reducing internal strain).

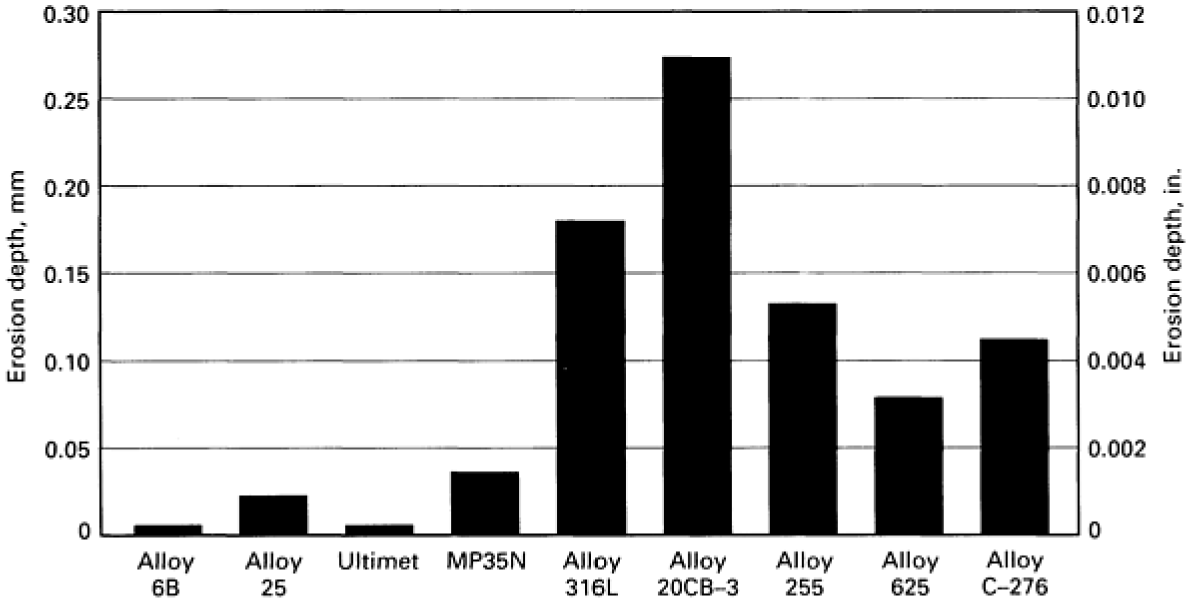


Fig. 8 Vibratory cavitation erosion test (ASTM G 32) results to relate cobalt-base wrought alloys with comparable alloys. Test parameters: test temperature, 16 °C (61 °F); test medium, distilled water; frequency, 20 kHz; amplitude, 0.05 mm (0.002 in.). All samples were solution annealed with the exception of MP35N alloy, which was tested in the work-strengthened condition.

The effect of carbon content (hence carbide volume fraction, and, indirectly, hardness) upon cavitation erosion resistance was studied in Ref 3. In the range studied (0.12 to 1.35 wt%), carbon content was found to have little effect above about 0.3 wt%. Below this level, cavitation erosion resistance was found to decrease with decreasing carbon content, although these workers neither addressed the issue of solubility limit nor the effect of carbon within the soluble range. It is interesting to note that the cavitation erosion resistance of Ultimet alloy (the carbon content of which is within the soluble range) is similar to that of Alloy 6B (at 1.0 wt% C).

Slurry Erosion. The slurry erosion properties of Alloy 6B, Alloy 25, and Ultimet alloy are presented in Fig. 9, together with data for commonly used stainless steels and nickel-base alloys. From these data, it is evident that cobalt as an alloy base can be beneficial, although the performance of Alloy 25 in this test (the details of which are described in Ref 12) was poor relative to most of the other alloys tested. It is also evident that carbide precipitates within the microstructure may be beneficial (comparing the erosion depths for Alloy 6B and Ultimet alloy).

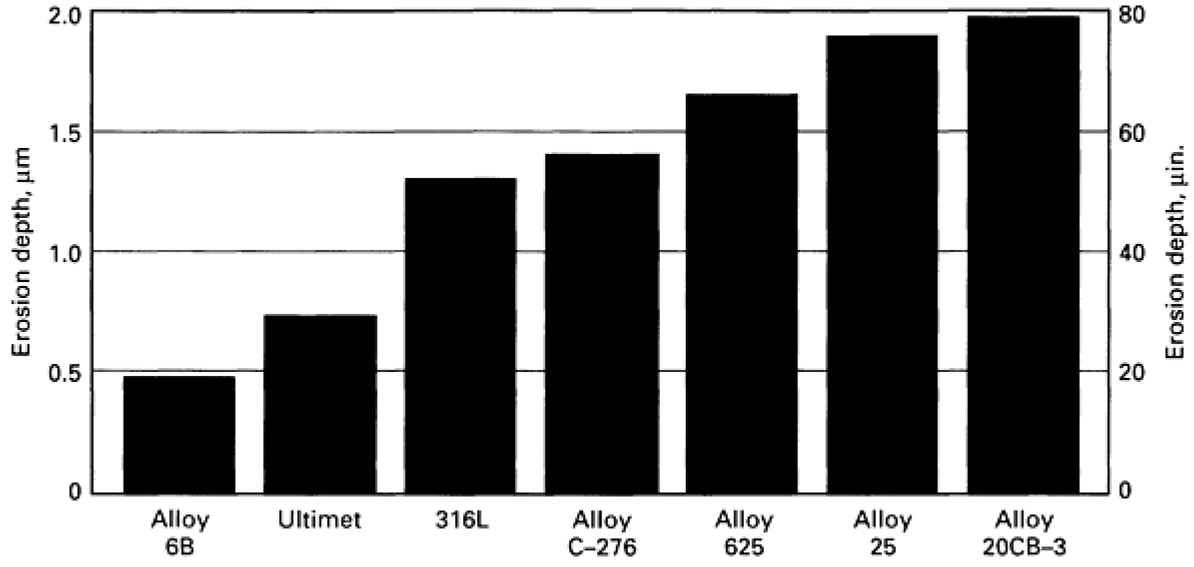


Fig. 9 Slurry erosion pot test data to relate cobalt-base wrought alloys with comparable alloys. Test parameters: test temperature, 20 °C (70 °F); test medium, 80 μm (0.003 in.) mean diameter alumina in tap water; particle loading, 0.12 kg/L; particle velocity, 5 m/s (16 ft/s); impact angle, 30°

These results are somewhat consistent with previous slurry erosion test data (Ref 13) on cobalt-base hardfacing alloys, albeit in a sand/water slurry, as opposed to the alumina/water medium used to generate the values in Fig. 9. In the previous work, a wide spectrum of engineering materials was tested, and, of these, Alloy 6 (the cast counterpart of Alloy 6B) was among the most resistant, wearing at only 5.5% of the rate of carbon steel. The results in Ref 13 indicated that the cast cobalt-base alloys of higher carbon content (and higher hardness) were much less resistant to slurry erosion, suggesting that the relationship between carbide volume fraction and slurry erosion resistance is far from linear for the cobalt-base alloys.

Sliding Wear

Of the three major types of wear, sliding is perhaps the most complex, not in concept, but in the way different materials respond to sliding conditions. Sliding wear is a possibility whenever two surfaces are forced together and moved relative to one another. The chances of damage are markedly increased if the two surfaces are metallic in nature, and if there is little or no lubrication present.

Galling. The outstanding resistance of the cobalt-base wrought alloys to galling, under self-mated conditions, is apparent from the pin-block test data presented in Fig. 10. Unfortunately, the test conditions (ten strokes through an arc of 120°) were too severe for the standard (austenitic and ferritic/austenitic) stainless steels. Even Nitronic 60, a type 200 stainless steel known for its resistance to galling, exhibited considerable damage (change in surface roughness) at the test load employed.

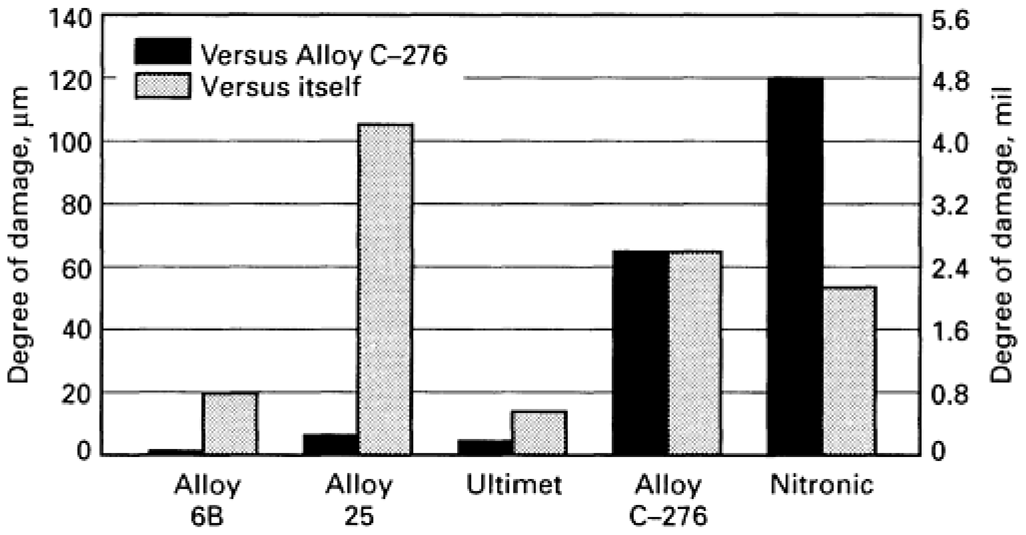


Fig. 10 Comparison of galling test data for cobalt-base wrought alloys with other selected alloys. Pin-on-block test parameters: test temperature, 20 °C (70 °F); number of strokes, 10 strokes through 120° arc; 2722 kg (6000 lb)

As may be deduced from Fig. 10, the coupling of dissimilar materials in a sliding system complicates the issue of wear performance, possibly because the interfacial bonding characteristics are difficult to predict. Nickel in Ultimet alloy, for example, is detrimental under self-mated conditions, but beneficial when sliding against alloy C-276 (Fig. 11).

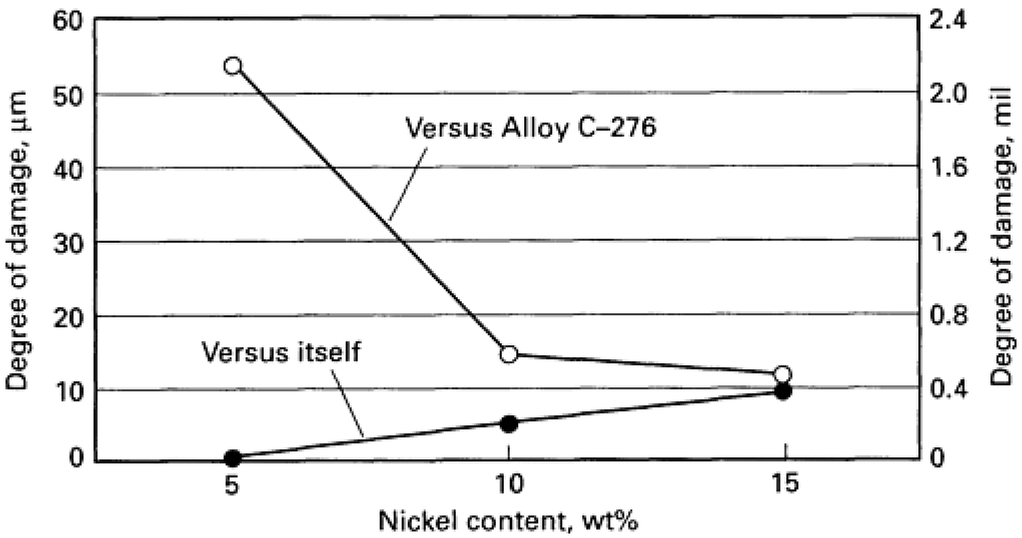


Fig. 11 Effect of nickel content on the antigalling properties of Ultimet alloy. Pin-on-block test parameters; test temperature, 20 °C (70 °F); number of strokes, 10 strokes through 120° arc; load, 2722 kg (6000 lb)

Oxide Control. Although elevated-temperature sliding wear data for the cobalt-base wrought alloys are not available, work with cobalt-base castings and weld overlays (Ref 14, 15) indicates that maximum wear rates occur at around 250 °C (480 °F) under a variety of sliding conditions and that above this temperature, oxide control of the sliding wear process is prevalent. Both Ref 14 and 15 discuss the advantages of cobalt as an alloy base over nickel (relative to sliding wear at low temperatures.)

The unlubricated dynamic coefficient of friction was found to vary with time in both of these studies. In Ref 14, the coefficient was found to vary between 0.4 and 1.4 for alloy 31 (self-mated) at room temperature (the value increasing with test time). At high temperatures, friction coefficient values appear to be more stable, a short, running-in period. At 800 °C (1470 °F), for example, Alloy 31 (self-mated) exhibited a value of between 0.3 and 0.4 after 15 min, up to completion of the test at 67 h.

Mechanical Properties

The mechanical properties of the cobalt-base wrought alloys are summarized in Vol 2 of the *ASM Handbook* (1990). Because they may be of some relevance to hot abrasion (low stress), recently generated hot hardness values are presented in Table 3 for Alloy 6B, Alloy 25, Alloy 188, and Ultimet alloy. It is significant that, despite the considerable spread in their room temperature hardness values, these alloys possess nearly equal resistance to deformation at 871 °C (1600 °F) and 982 °C (1800 °F).

Table 3 Hot hardnesses of cobalt-base wrought alloys

Temperature		Hardness for indicated alloy, DPH[(a)]			
°C	°F	Alloy 6B	Alloy 25	Alloy 188	Ultimet
RT[(b)]	RT[(b)]	374	285	248	296
427	800	269	171	170	173
538	1000	247	160	159	162
649	1200	225	150	147	158
760	1400	153	134	129	134
871	1600	91	93	89	89
982	1800	55	52	49	50

Source: Ref 16

(a) DPH, diamond pyramid hardness

(b) RT, room temperature.

Wear-Related Applications

As may be deduced from the wear data presented in this article, the cobalt-base alloys provide high resistance to sliding wear, slurry erosion, and cavitation erosion (relative to the stainless steels, nickel-base superalloys, and high-nickel corrosion alloys). Alloy 6B and Alloy 6K also possess outstanding resistance to low-stress abrasion. Industrial applications of the alloys tend to reflect these general properties and their resistance to corrosion and high-temperature strength.

The alloys are used for sliding wear applications such as valves, both in the chemical processing industries and in the power industries (largely for steam control). Furthermore, Alloy 25 is used as a bearing material, and Alloy 6B is widely used to protect the nose sections of industrial chain saws. Ultimet alloy has been used for electrogalvanizing rolls, where both corrosion and sliding wear resistance are required.

The erosion properties of the cobalt-base alloys have led to their use in nozzles, valves (subjected to cavitation damage or slurries), pumps, and mixing blades. Alloy 6B is also used to protect steam turbine blades from liquid droplet erosion (it appears that the advantages of cobalt as an alloy base extend to this form of erosion also).

From an abrasion standpoint, notable applications of Alloy 6B and Alloy 6K include scraper knives and blades (for the chemical and food processing industries) and knives for the textile industries.

References

1. D.A. Woodford, Cavitation-Erosion-Induced Phase Transformations in Alloys, *Metall. Trans.,* Vol 3, 1972, p 1137
2. K.J. Bhansali and A.E. Miller, Role of Stacking Fault Energy on the Galling and Wear Behavior of a Cobalt Base Alloy, *Wear of Materials,* American Society of Mechanical Engineers, 1981, p 179
3. K.C. Antony and W.L. Silence, The Effect of Composition and Microstructure on Cavitation Erosion Resistance, *Proceedings of the 5th International Conference on Erosion by Solid and Liquid Impact,* Cambridge University Press, 1979, p 671
4. C.J. Heathcock, B.E. Protheroe, and A. Ball, The Influence of External Variables and Microstructure on the Cavitation Erosion of Materials, *Proceedings of the 5th International Conference on Erosion by Solid and Liquid Impact,* Cambridge University Press, 1979, p 631
5. C.T. Sims, Cobalt-Base Alloys, *The Superalloys,* John Wiley & Sons, 1972, p 145
6. W.L. Silence, Effect of Structure on Wear Resistance of Co-, Fe-, and Ni-Base Alloys, *Wear of Materials,* American Society of Mechanical Engineers, 1977, p 77
7. R.B.H. Herchenroeder, S.J. Matthews, J.W. Tackett, and S.T. Wlodek, Haynes Alloy No. 188, *Cobalt,* Vol 54, 1972, p 3
8. H.M. Tawancy, V.K. Ishwar, and B.E. Lewis, On the FCC to HCP Transformation in a Cobalt-Base Superalloy (Haynes Alloy No. 25), *J. Mater. Sci. Lett.,* Vol 5, 1986, p 337
9. "Metal Corrosion, Erosion, and Wear," G 76, *Annual Book of ASTM Standards,* Vol 3.02, ASTM, 1984, p 438-445
10. A. Ninham, The Effect of Mechanical Properties on Erosion, *Wear,* Vol 121, 1987, p 307
11. I.M. Hutchings, *The Erosion of Materials by Solid Particle Impact,* No. 10, MTI, 1983
12. A.V. Levy and S. MacAdam, "Erosion-Corrosion of Materials in Coal-Water Slurries," Report AP-5628, Electric Power Research Institute, 1988
13. W.A. Stauffer, Wear of Metals by Sand Erosion, *Met. Prog.,* Jan 1956, p 102
14. F.H. Stott, C.W. Stevenson, and G.C. Wood, Friction and Wear Properties of Stellite 31 at Temperatures

from 293 to 1073 K, *Met. Technol.,* Feb 1977, p 66
15. P. Crook and C.C. Li, The Elevated Temperature Metal-to-Metal Wear Behavior of Selected Hardfacing Alloys, *Wear of Materials,* American Society of Mechanical Engineers, 1983, p 272
16. S.K. Srivastava, Haynes International Database

Friction and Wear of Ordered Intermetallic Alloys of Ni_3Al

Peter J. Blau, Metals and Ceramics Division, Oak Ridge National Laboratory

Introduction

ALLOYS based on the ordered intermetallic compound Ni_3Al are a relatively new development; however, the compound has long been alternatively known as the γ' phase in superalloys. Owing to the alloys' relatively recent introduction, most of the information available on the tribological properties of these alloys has been derived from laboratory tests and not from engineering applications. However, a number of laboratory studies have indicated that nickel aluminide alloys have significant potential in wear-critical applications, especially in cavitation erosion and in sliding wear at temperatures between ~400 and 650 °C (~750 and 1200 °F). In this article, the compositions, physical properties, mechanical properties, and limited friction and wear data for several of the leading compositions of nickel aluminide alloys are provided. The wear data include cavitation wear, solid-particle erosive wear, abrasive wear, and sliding wear.

Composition and Structure of Nickel Aluminide Alloys

Commercial versions of nickel aluminide alloys have recently been developed. In the United States, Oak Ridge National Laboratory (ORNL) has been a major developer; consequently, its numbering system has been adopted for use in this article. Unless otherwise cited, property data were obtained from the Metals and Ceramics Division of ORNL (Ref 1). The compositions of several nickel aluminide alloys are given in Table 1. A more detailed discussion of the structure and properties of ordered intermetallic alloys, including nickel aluminide alloys, is given by Liu *et al.* in the article "Ordered Intermetallics" in *Properties and Selection: Nonferrous Alloys and Special-Purpose Materials,* Volume 2 of the *ASM Handbook,* formerly 10th Edition *Metals Handbook* (Ref 2).

Table 1 Nominal compositions of selected nickel aluminide alloys

Alloy[a]	Composition, wt%						
	Al	**Cr**	**Fe**	**Zr**	**Mo**	**B**	**Ni**
IC-50	11.3	. . .	. . .	0.6	. . .	0.02	bal
IC-74M	12.4	. . .	. . .	. . .	. . .	0.05	bal
IC-218	8.5	7.8	. . .	0.8	. . .	0.02	bal
IC-218 LZr	8.7	8.1	. . .	0.2	. . .	0.02	bal
IC-221	8.5	7.8	. . .	1.7	. . .	0.02	bal
IC-357	9.5	7.0	11.2	0.4	1.3	0.02	bal
IC-396M	8.0	7.7	. . .	0.8	3.0	0.01	bal

(a) Designations used by Oak Ridge National Laboratory, Oak Ridge, TN

During the later part of the 1970s, it was found that small boron additions were critical for achieving reasonable levels of ductility in the alloys. Boron is thought to increase grain boundary cohesiveness, thereby reducing the tenacity for brittle intergranular fracture. Other alloying additions were made to improve strength, castability, hot workability, and corrosion resistance.

Ordered nickel aluminide (Ni_3Al) alloys are based on an $L1_2$ crystal structure. The unit cell consists of a face-centered cubic (fcc) arrangement in which the aluminum atoms occupy the corner positions and the nickel atoms preferentially occupy the face-centered positions. The fact that atoms occupy preferred positions in the crystal structure imparts unusual mechanical properties to the nickel aluminide alloys. These properties will be described in the section "Mechanical Properties" in this article. The presence of superdislocations, their dissociation into partials, and their associated mechanisms of cross-slip affect the temperature versus strength behavior of Ni_3Al alloys.

Nickel aluminide alloys are still under development and commercial evaluation. It is likely that new compositions will become available. However, the commercialization of the alloys will be strongly driven by the price of nickel and by the presence of specific applications for which nickel aluminide alloys may be the only performance-based choice.

Physical Properties

Little data exist on the physical properties of the whole series of nickel aluminide alloys; however, the information in Table 2 summarizes the effect of small changes in stoichiometry of Ni_3Al on several physical properties of the material (Ref 3). Data showing the effect of temperature on electrical resistivity, thermal conductivity, and the mean coefficient of thermal expansion are also given in Table 2.

Table 2 Physical properties of selected nickel-aluminum binary alloys

Data are for arc-melted, homogenized, and annealed specimens produced from high-purity nickel and aluminum stock.

Temperature		Electrical resistivity		Thermal conductivity		Mean coefficient of thermal expansion	
°C	°F	($\Omega \cdot$ m) $\times 10^{-8}$	m$\Omega \cdot$ in.	W/m · K	Btu · in./ft² · h · °F	(m/m · K) $\times 10^{-6}$	in./in. · °F
24.0 at% aluminum content[(a)]							
27	81	51.3	20.1	21.4	148	. . .	. . .
427	801	72.6	28.5	30.1	208	. . .	. . .
727	1341	84.2	33.0	32.7	227	. . .	. . .
27-327	81-621	. . .	. . .	. . .	. . .	12.5	6.9
427-727	801-1341	. . .	. . .	. . .	. . .	14.8	8.2
727-1027	1341-1881	. . .	. . .	. . .	. . .	16.8	9.3
25.0 at% aluminum content[(b)]							
27	81	32.6	12.8	28.9	200	. . .	. . .
427	801	55.3	21.7	37.1	257	. . .	. . .
727	1341	73.0	28.6	36.1	250	. . .	. . .
27-327	81-621	. . .	. . .	. . .	. . .	12.5	6.9
427-727	801-1341	. . .	. . .	. . .	. . .	14.6	8.1
727-1027	1341-1881	. . .	. . .	. . .	. . .	16.8	9.3

(a) Density, 7.49 g/cm³ (0.270 lb/in.³) at 22 °C (72 °F).

(b) Density, 7.43 g/cm³ (0.268 lb/in.³) at 22 °C (72 °F)

Mechanical Properties

Among the most striking properties of nickel aluminide alloys, and that which captured initial interest among metal producers, is their rising yield strength with rising temperature. Figure 1 shows that the yield strength of four nickel aluminide alloys tends to rise to a maximum in the temperature range of ~400 to 650 °C (~750 to 1200 °F). Above this temperature range, the yield strength declines. The effect of cold working on the yield strength of IC-50 alloy is shown in Fig. 2. Up to test temperatures of ~600 °C (~1110 °F), cold working had an effect on the yield strength, but at a test temperature of 800 °C (1470 °F), there was no significant enhancement of yield strength.

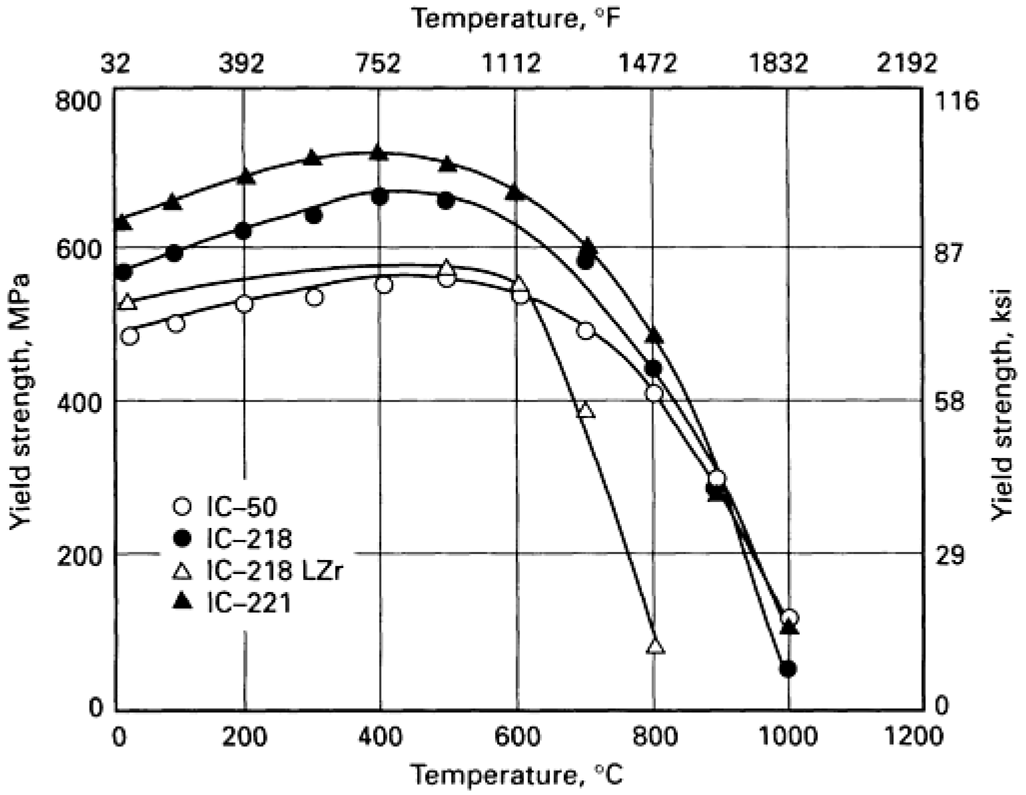


Fig. 1 Variation of yield strength with test temperature for selected nickel aluminide alloys. Strain rate, 0.5 mm/mm/min

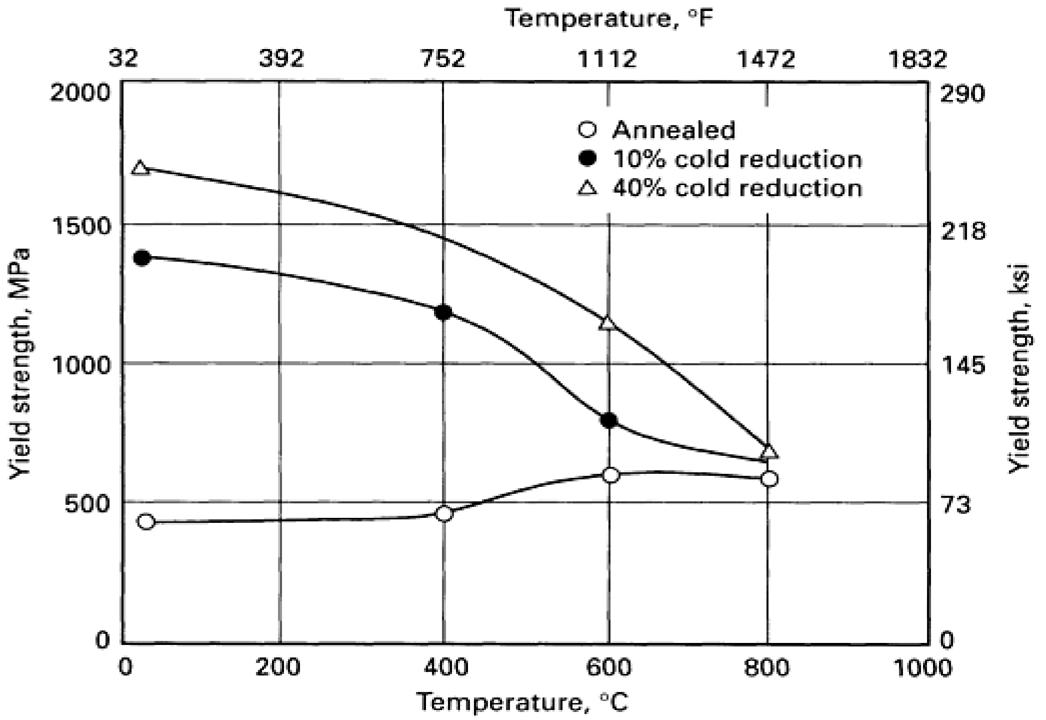


Fig. 2 Plot of yield strength versus test temperature of IC-50 nickel aluminide alloy as a function of cold working

The elastic (Young's) modulus of four nickel aluminide alloys is given in Table 3. Data on the effects of temperature on the elastic modulus of three of the alloys are also provided.

Table 3 Elastic modulus of four nickel aluminide alloys at selected temperatures

Temperature		Elastic modulus of Indicated alloy							
		IC-50		**IC-218**		**IC-218 LZr**		**IC-221**	
°C	**°F**	**GPa**	**10^6 psi**	**GPa**	**10^6 psi**	**GPa**	**10^6 psi**	**GPa**	**10^6 psi**
25	77	173	25	228	33	186	27	214	31
200	329	. . .	. . .	. . .	. . .	186	27	. . .	. . .
400	752	152	22	166	24	166	24	. . .	. . .
600	1112	145	21	. . .	. . .	152	22	. . .	. . .
800	1472	117	17	173	25	117	17	. . .	. . .

As with most metal alloys, the manner of fabrication and treatment affects the hardness of nickel aluminide alloys. Knoop microindentation hardness data for six nickel aluminide alloys in different forms are given in Table 4. All table data were obtained with a 100 gf (0.981 N) load. As indicated by IC-50 data in Table 4, the hardness of the alloy was observed to vary by as much as a factor of 1.7 from one form to another. Knoop microindentation hardness numbers well above 6.0 GPa (612 kg/mm^2) have been recorded on cross sections of heavily deformed machining chips of nickel aluminide alloy IC-218 LZr (Ref 4). Similar values are observed in the highly deformed, subsurface layers produced by unlubricated sliding contact of this alloy with tungsten carbide at room temperature (RT). The microindentation hardnesses of several of the counterface materials--used in the sliding experiments to be discussed in the section "Sliding Friction and Wear Data" in this article--are also given in Table 4.

Table 4 Knoop microindentation hardness of nickel aluminide alloys and selected slider experiment counterface materials

Material	Condition	Hardness, HK[(a)]	
		Average GPa[(b)]	**Limits ± GPa**
Nickel aluminide alloys			
IC-50	Small cast ingot	1.96	0.05
	Cross section of a 1 mm (0.04 in.) cast strip	2.91	0.12
	Surface of 6.4 mm (0.25 in) cast rod	3.39	0.28
IC-74M	Rolled and annealed strip	2.37	0.28
IC-218	Dendrite arm in a small cast ingot	2.65	0.12
IC-218 LZr	Small cast ingot	3.84	0.60
IC-221	Dendrite arm in a small cast ingot	2.73	0.20
IC-396M	Small cast ingot	2.80	0.16
	Surface of 6.4 mm (0.25 in.) cast rod	3.93	0.23
Counter materials[(c)]			
Counterface material[(a)] 52100 bearing steel	Surface of a 9.5 mm (0.375 in.) bearing ball	10.6	1.0
NBD 100	Surface of a 9.5 mm (0.375 in.) silicon nitride ball	28.0	3.0

(a) Load, 0.98 N (100 gf.)

(b) Data are the average of five reported values. To convert GPa to kg/mm^2, multiply GPa by 101.94.

(c) Data for wear test specimens used as sliders

Nickel Aluminide Wear Data

Nickel aluminide wear data, as presented here, fall into three main categories:

- Erosive wear (which includes cavitation erosion and solid-particle erosion

- Abrasive wear
- Sliding wear

Cavitation Erosion

A study of the cavitation erosion behavior of several nickel aluminide alloys in pressurized water was conducted at the Norris Engineering Laboratory of the Tennessee Valley Authority using a specially designed cavitating jet apparatus (Ref 5, 6). High-pressure water was supplied by a positive-displacement pump (rated for a flow of 0.19 L/s, or 3.0 gal/min, at 69.0 MPa, or 10 ksi) to an adjustable nozzle. Data for a number of alloys applied to test specimen surfaces or as bulk specimens subjected to 41.6 MPa (6 ksi) operating pressure are given in Table 5. The data in Table 5 show that, of all the materials tested, three nickel aluminide alloys were the most resistant to cavitation erosion.

Table 5 Comparison of jet cavitation erosion data for three nickel aluminide alloys and other selected alloys relative to cold-rolled carbon steel

Operating pressure was 40 MPa (6 ksi).

Alloy	Cavitation erosion rate	
	Weight loss, mg/h	**Relative rate**
Carbon steel (cold-rolled bar)	32.3	1.0
Type 304 stainless steel plate	33.2	1.1
Type 308 stainless steel overlay (stick weld)	31.6	1.0
Nitronic 60 bar stock	12.8	0.4
Stellite 21 (stick)	7.3	0.2
IC-50 nickel aluminide (plasma weld overlay)	7.0	0.2
Stoody Deloro HGl (4 mm, or $\frac{5}{32}$ in., stick)	6.1	0.2
IC-50 as-formed	4.1	0.1
IC-218 nickel aluminide	2.9	0.09
IC-221 nickel aluminide	2.5	0.08
IC-50 nickel aluminide (cold worked)	1.4	0.04

Source: Ref 7

Solid-Particle Erosion

Sahoo *et al.* (Ref 8) conducted a study of the solid-particle erosion resistance of nickel aluminide alloys IC-50 and IC-218, in both annealed and cold-rolled conditions, compared with type 304H stainless steel. The alloys were cut into 25 × 30 mm (0.98 × 1.2 in.) coupons and finished with 600-grit abrasive prior to testing. Silicon carbide grains averaging 180 μm (0.0072 in.) diameter were used as the erodent at a measured nozzle velocity of 25 m/s (82 ft/s) and fed into the preheated gas stream at 10g/min (0.22 lb/min). The sample chamber was also heated. A carousel-type specimen holder permitted up to four specimens to be tested sequentially. After being exposed to 100 g (0.2 lb) of erodent, the specimens were each weighed to determine the erosion rate. Table 6 shows the effects of carrier gas, temperature, impingement angle, and alloy condition on the erosion rate. For nickel aluminide alloys that were eroded at 600 °C (1110 °F), there was a slight reduction in the erosion rate for the air atmosphere compared to the nitrogen atmosphere in the case of the cold-rolled samples. The particle impingement angle did have a marked effect on the erosion rate. As is characteristic of metals, the 30° impact angle gave higher rates than the 90° impact angle. Temperature of erosion had very little effect on the erosion rates of nickel aluminide, although it appeared to slightly increase the low-angle erosion rate of the steel and decrease the normal incidence erosion rate of the steel. Overall, there was not much difference in the erosion rates of nickel aluminide alloys when compared to type 304H stainless steel under the same conditions. As with the cavitation experiments discussed previously in the section "Cavitation Erosion" in this article, these tests represent only a small number of conditions, and the erosion behavior of nickel aluminide alloys may vary considerably from the present results under other conditions.

Table 6 Solid-particle erosion data for IC-50 and IC-218 twin-roll cast nickel aluminide alloys relative to type 304H stainless steel when using silicon carbide as the erodent

Test conditions[a]				Erosion rate as a function of indicated alloy and alloy processing[b]				
Carrier gas	Gas stream temperature		Impingement angle	IC-50		IC-218		Type 304H stainless steel
	°C	°F		Annealed	Cold rolled	Annealed	Cold rolled	
Nitrogen	25	75	30°	4.0	4.2	4.0	4.3	3.7
			90°	2.4	2.2	2.4	2.5	2.8
	600	1110	30°	3.8	4.3	3.9	4.3	4.2
			90°	2.3	2.5	2.2	2.6	2.2
Air	600	1110	30°	3.8	4.0	3.8	3.9	4.2

Source: Ref 8

(a) Test parameters: particle size, 180 μm (0.0072 in.) diameter; gas stream flow, 10 g/min (0.02 lb/min); nozzle velocity, 25 m/s (82 ft/s).

(b) Erosion rate is figured as (grams material loss/100 g erodent) × 10^{-4}.

Abrasive Wear

Two recent studies have involved the abrasive wear behavior of nickel aluminide alloys (Ref 6, 9). Marquardt and Wert (Ref 9) used 400-grit silicon carbide abrasive paper affixed to a rotating flat disk to test the two-body abrasion behavior of annealed IC-15 (12.7 wt% Al, 0.05 wt% B, and 87.3 wt% Ni) pin specimens. The sliding alloy pin was precessed during the experiment such that new abrasive was continuously exposed to the slider. Each test at the stated normal force was repeated three times for a total sliding distance of 90 m (300 ft). A summary of the results is given in Table 7. These results indicate that under the stated conditions, IC-15 was slightly worse in abrasion resistance than type 316 stainless steel, a material not generally recommended for abrasive wear conditions.

Table 7 Two-body abrasive wear data for IC-15 nickel aluminide alloy relative to selected steel alloys

Normal force		Abrasive wear rates for indicated alloy, mg/m			
N	kgf	52100 bearing steel	Type 304L stainless steel	Type 316 stainless steel	IC-15
1.96	0.196	0.127	0.171	0.171	0.190
4.91	0.491	0.297	0.434	0.387	0.482

Source: Ref 9

Johnson *et al.* (Ref 6) studied the three-body abrasive wear of three nickel aluminide alloys: IC-50, IC-218, and IC-221. They used the ASTM standard test method G 65, which involves sliding of test coupons against a moving stream of sand particles held against the specimen by a rotating rubber-faced wheel. The parameters used include 50 N (5 kgf) load, wheel diameter of 229 mm (9.02 in.) rotation at 230 rev/min, 60-mesh dry sand flow rate of 4.2 cm^3/s (0.26 $in.^3/s$) and a test duration of 540 s. Of the three alloys tested, IC-50 (the basic Ni-Al-B composition) exhibited the best abrasion resistance.

Sliding Friction and Wear Data

Only limited laboratory data on the sliding friction and wear data is available for nickel aluminide alloys. Investigations were done at Vanderbilt University (Ref 9), The Ohio State University (Ref 10), and at Oak Ridge National Laboratory (Ref 11, 12, 13). Marquardt and Wert (Ref 9) compared the sliding wear of IC-15 with 52100 bearing steel, type 304L stainless steel, and type 316 stainless steel in self-mated, pin-on-disk sliding wear tests at room temperature. Relative to the wear rate of the 52100 steel pin, which wore the least, the wear rates of the other materials were: type 316 stainless steel, 19.3; type 304L stainless steel, 27.3; and IC-15, 38.3. Therefore, the nickel aluminide alloy experienced the worst wear of all four materials at room temperature.

Bonda and Rigney (Ref 10) investigated the sliding wear of Ni_3Al with and without boron and IC-218 in both air and vacuum (1 mPa, or 0.15 × 10^{-6} psi) at room temperature and at 400 °C (750 °F). The basic alloys were produced by thermal spraying and the IC-218 was produced by hot extrusion of the alloy powder. The test geometry was the edge of a small disk slidingon the surface of a larger, vertically oriented disk. The vertical disk orientation was used so that debris could be collected below the contact. Materials were slid against themselves. Normal load was 14.7 N (1.5 kgf), and sliding speeds were 15 to 19 mm/s (0.59 to 0.75 in./s). Sliding was smoother in air than in vacuum for all materials at room temperature, but the wear rates were much higher in air. The effect of temperature was found to be small in a vacuum, but large in air, suggesting the importance of oxygen in the wear process. The wear behavior of the basic alloy was not affected by the presence of boron under all testing conditions. In air and at both room temperature and 400 °C (750 °F), the disks of IC-218 alloy wore the most of all three materials.

The remainder of the friction and wear test data presented here where obtained at Oak Ridge National Laboratory. Two types of sliding experiments are described:

- Unidirectional pin-on-disk experiments
- Reciprocating-sphere or cylinder-on flat experiments

Unidirectional, Unlubricated Sliding Results. Unidirectional sliding friction and wear data were obtained using the equipment illustrated in Fig. 3 and 4. In all experiments, the pin specimen was a 9.53 mm ($\frac{3}{8}$ in.) diameter ball of the specified material held fixed in the holder. The natural frequency of the apparatus under load was about 20 Hz. All tests were conducted in air at typical relative humidity levels ranging from 40 to 70%. The pin was deadweight loaded to provide the reported normal forces. The sliding velocity was 0.1 m/s (0.3 ft/s) except where otherwise specified. Test duration was either 10,000 disk revolutions or 1 km (3000 ft) sliding distance. In the current case and considering the range of wear track radii involved in these experiments, these durations are approximately the same (±15%). Disk specimens were metallographically polished and cleaned in acetone and methanol before testing. Spheres were also solvent cleaned before testing. Friction values reported are the nominal values obtained after running-in transients had subsided.

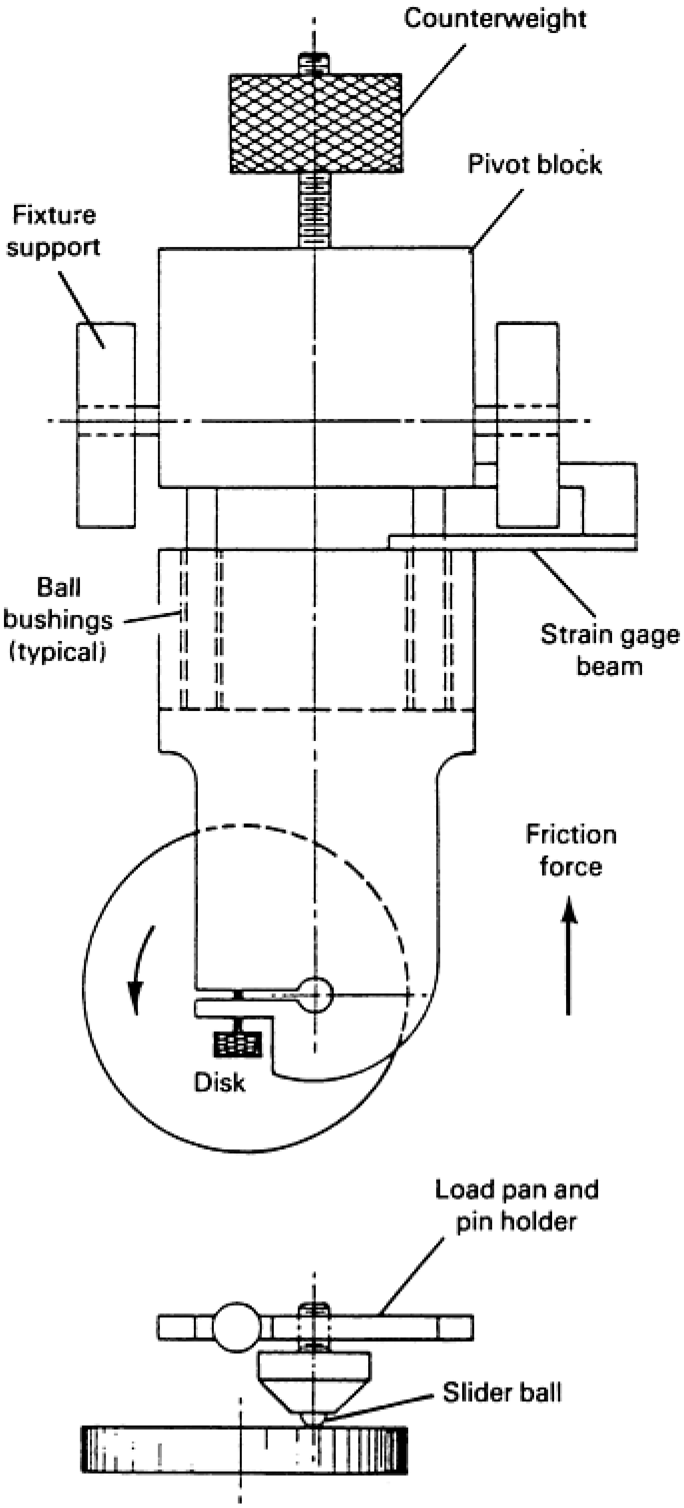


Fig. 3 Pin-on-disk test apparatus used to obtain unidirectional sliding friction and wear data in air at room temperature

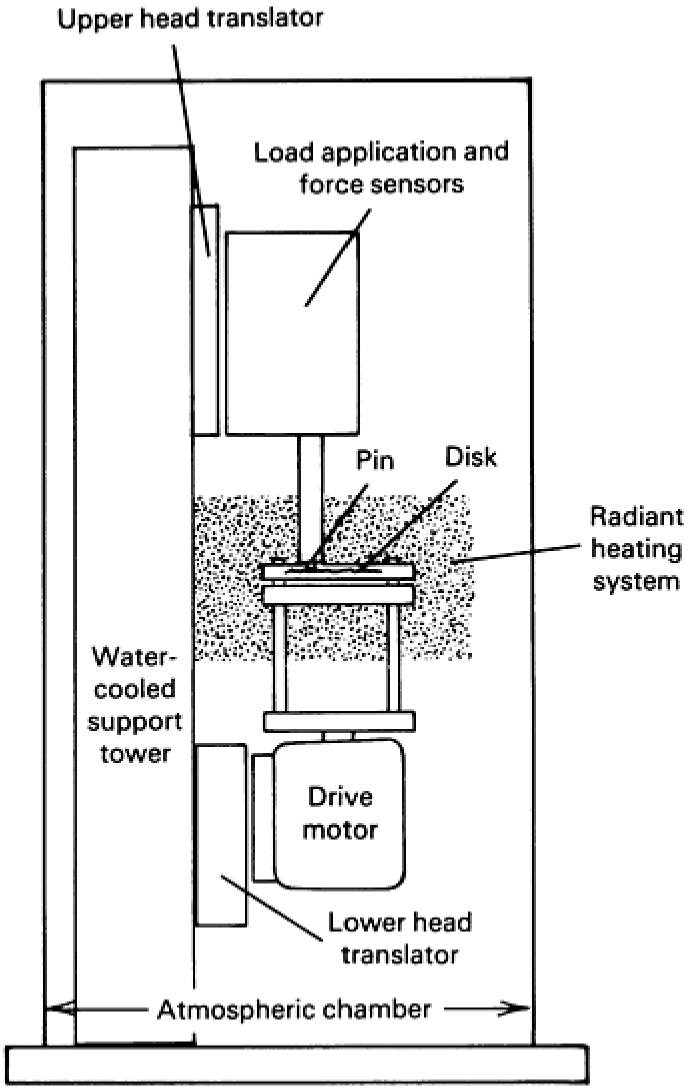


Fig. 4 Pin-on-disk test device used to measure unidirectional sliding friction and wear in a vacuum at both room temperature and elevated temperatures

Wear rates of the pin specimens were obtained by photographing each wear scar, measuring its diameter, and computing the volume removed based on the assumption of a flat-based, spherical cap. Wear rates of the flat, disk specimens were obtained from stylus tracings across the wear tracks at at least three locations and then by multiplying the average, digitized cross-sectional area of the track by the track length (computed from the diameter of the track measured at its center). All wear rates have been expressed in the traditional "wear factor" notation of volume lost normalized by the total sliding distance and by the applied load (that is, $mm^3/N \cdot m$). It is recognized that the contact by the pin tip is approximately continuous and that the contact by any point on the disk track is intermittent. Therefore, normalizing the disk wear volume by the total sliding distance experienced by the pin is fundamentally incorrect. However, a great quantity of historical pin-on-disk results have been reported on this basis, and in order to compare the current results with these values, the former method of wear volume normalization was retained.

Pin-on-disk friction and wear data for five nickel aluminide alloys against several metals and ceramics (tested both at room temperature and at elevated at temperatures) are given in Table 8. The friction coefficients ad wear rates of the various nickel aluminide combinations at room temperature are very similar to those steel on steel and do not represent any significant tribological advantages in this area. Studies of combined wear and corrosion of these alloys have yet to be done.

Table 8 Pin-on-disk friction and wear data for several nickel aluminide alloys tested in air at selected temperatures with 10 N (1 kgf) load

Disk alloy	Test temperature		Slider ball material[a]	Nozzle velocity		Friction coefficient[b]	Wear factor[c] (mm^3/N · m) × 10^{-7}	
	°C	**°F**		**m/s**	**ft/s**		**Pin[d]**	**Disk**
IC-50	RT	RT	Al_2O_3	0.1	0.3	0.52	1.6	479
	RT	RT	52100 bearing steel	0.1	0.3	0.70	NM[f]	1140
	RT	RT	M50 tool steel	0.1	0.3	0.68	14.0	1220
	400	750	Al_2O_3	0.10	0.33	0.43	19.9	369
	650	1200	Al_2O_3	0.10	0.33	0.38	NM[f]	645
IC-74M	RT	RT	Al_2O_3	0.1	0.3	0.60[e]	1.6	400
	400	750	Al_2O_3	0.10	0.33	0.65	18.0	1.12
	650	1200	Al_2O_3	0.10	0.33	0.60[e]	4.0	0.63
IC-218	RT	RT	Al_2O_3	0.1	0.3	0.47	NM	2511
	650	1200	Al_2O_3	0.10	0.33	0.43	. . .	794.0
IC-218 LZr	RT	RT	WC	0.64	2.1	0.55	NM[f]	1670
	RT	RT	WC	0.91	3.0	0.59	NM	3190
	650	1200	WC	0.64	2.1	0.70	16.8	34.6
	650	1200	WC	0.91	3.0	0.70	14.9	25.3
IC-221	RT	RT	Al_2O_3	0.1	0.3	0.60	NM	794

(a) Ball diameter, 9.53 mm (0.375 in.).

(b) Nominal value at steady state averaged for at least two tests. Actual value may vary by as much as 25%.

(c) Average value for at least two tests. Test-to-test values may vary by as much as 33%.

(d) NM, not measurable.

(e) Load, 0.98 N.

(f) Not measurable due to material adhering to the pin tip

The friction coefficients of the sliding couples were not substantially better at high temperature than the values obtained at room temperature; however, the high-temperature wear rates of the IC-218 LZr against tungsten carbide and the IC-74M against alumina (Fig. 5) were much better than those at room temperature. This behavior and the results from the referenced studies suggest that the nickel aluminide alloys have the potential to serve as sliding partners for ceramics in the temperature range of 400 to 650 °C (750 to 1200 °F). More work on modifying the thermomechanical processing of nickel aluminide alloys for peak wear resistance is needed to maximize the improved wear resistance that these alloys have shown at elevated temperatures.

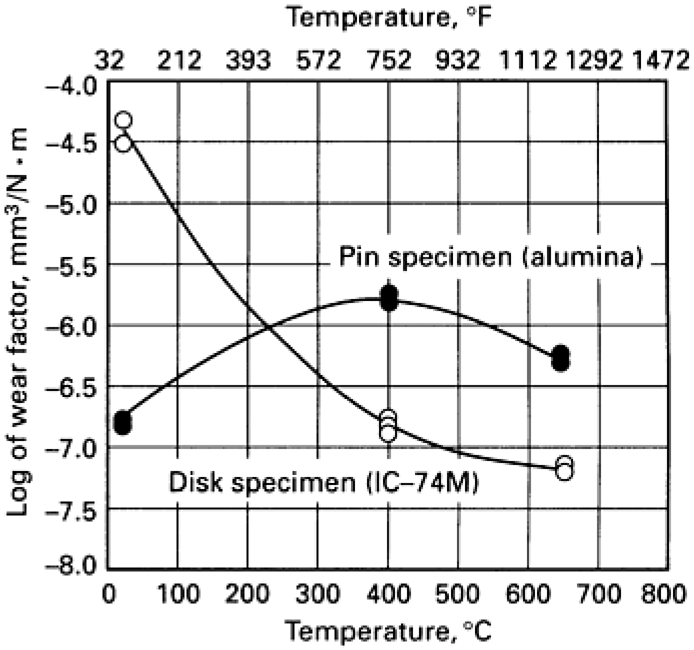


Fig. 5 Wear data for IC-74M disk against alumina pin in air at selected temperatures (room temperature; 400 °C, or 750 °F; and 650 °C, or 1200 °F). Pin-on-disk apparatus parameters: load, 1.0 N (0.1 kgf); sliding speed, 0.1 m/s (0.3 ft/s); wheel travel, 10,000 rotations

Reciprocating Sliding Friction and Wear Tests. Reciprocating (back-and-forth) testing of various combinations of nickel aluminide alloys with other materials were conducted on a commercially manufactured Cameron-Plint TE-77 high-frequently machine. As with the unidirectional tests, the slider was a 9.53 mm (0.375 in.) diameter ball clamped in the device. The frequency of oscillation was set at 10.0 Hz, the stroke length was 10.0 mm (0.394 in.), and the normal load was 25.0 N (2.55 kgf). Both unlubricated and lubricated tests were performed in room-temperature air. The duration of the test was 600 s (120 m, or 390 ft, sliding distance). Wear volume for the slider specimens was determined in the manner used for unidirectional tests described in the section "Unidirectional, Unlubricated Sliding Results" in this article. Flat-specimen wear volumes were obtained by three stylus traces across each scar, but the wear factor for the flat specimen was reported in terms of $mm^3/N \cdot cycle$ instead of $mm^3/N \cdot m$ to account for the fatigue aspects of this type of test.

Table 9 lists the unlubricated friction and wear data for two nickel aluminide alloys sliding against three different materials at room temperature at an average velocity of 0.2 m/s (0.7 ft/s). The bearing steel/nickel aluminide combination produced the most wear in both alloys, and the silicon nitride/nickel aluminide combination was best. Table 9 also lists the friction and wear for the same material combinations in the presence of SDL-1, a formulated SAE 15W-30 commercial lubricant. The steel/nickel aluminide combination was benefited most by the presence of the lubricant. The alumina pins wore the least under the lubricated conditions, and they produced the highest wear of the nickel aluminide. Therefore, of the two ceramics (silicon nitride and alumina), the silicon nitride/nickel aluminide combination produced the lowest system (total) wear, especially in the case of IC-396M.

Table 9 Reciprocating ball-on-flat friction and wear data for two nickel aluminide alloys at room temperature in air

Test parameters: load, 25.0 N (2.55 kgf); oscillation frequency, 10.0 Hz; stroke length, 10.0 mm (0.394 in.); slider bar diameter, 9.53 mm (0.375 in.)

Parameter	Slider ball material					
	Al_2O_3, flat material		**Si_3N_4, flat material**		**52100 bearing steel, flat material**	
	IC-50	**IC-396M**	**IC-50**	**IC-396M**	**IC-50**	**IC-396M**
Unlubricated						
Friction coefficient[(a)]	0.80	0.80	0.72	0.80	1.13	1.08

Wear factor [(mm³/N · m) × 10⁻⁷][(b)]:						
Ball	560	590	140	140	660	1000
Flat	78	250	75	80	290	380
Lubricated (SDL-1 commercial lubricant 15W-30)						
Friction coefficient	0.07	0.13	0.06	0.10	0.06	0.09
Wear factor[(b)], (mm³/N · m) × 10⁻⁷						
Ball	NM[(c)(d)]	NM[(c)]	0.56	0.26	0.18	0.11
Flat	4.7	3.3	4.5	1.9	0.6	0.39

(a) Root-mean-square at steady state averaged for at least two tests. Actual value may vary by as much as 25%.

(b) Average value for at least two tests. Test-to-test values may vary by as much as 50%.

(c) NM, not measurable.

(d) Not measurable due to the absence of a discernible flat spot on the pin tip

Cylindrical rods, 6.0 mm (0.24 in.) in diameter, of IC-50 and IC-396M alloys were used for reciprocating tests with the cylinder-on-flat geometry. In each case, the long axis of the 10 mm (0.39 in.) long cylinders was fixed perpendicular to the sliding direction, and a rectangular scar was produced along the length of the cylinder. These rod materials were run against flats of the same materials. The hardness of the cylinders tended to be higher than that of the flats (see Table 4). As Table 10 data show, the flat specimens of IC-396M wore less than those of IC-50 at both test temperatures. Because the wear of the cylindrical specimens was similar in all tests, the total system wear, coupled with the slightly lower friction, makes IC-396M the best performer in these tests.

Table 10 Unlubricated reciprocating cylinder-on-flat friction and wear data for two self-mated nickel aluminide alloys

Parameters[(a)]	Test temperature			
	25 °C (75 °F)		**200 °C (390 °F)**	
	IC-50	**IC-396M**	**IC-50**	**IC-396M**
Friction coefficient[(b)]	0.99	0.80	0.98	0.79
Wear factor[(c)]				
Cylinder, (mm³/N · m) × 10⁻⁷	710	860	690	650

(a) Load, 25.0 N (2.55 kgf); oscillation frequency, 10.0 Hz; 10.0 mm (0.394 in.) stroke length; 6.0 mm (0.236 in.) diameter cylindrical sliders.

(b) Root mean square at steady state averaged for at least two tests. Actual value may vary by as much as 25%.

(c) Average value for at least two tests. Test-to-test values may vary by as much as 50%.

(d) NM, not measured

Sliding Wear Properties of Nickel Aluminide Alloys. Based on limited test data for nickel aluminide alloys in laboratory conditions, there appears to be a potential for these alloys in high-temperature friction and wear applications. However, a great deal more development work and field trials, aimed at specific wear applications, must be conducted before specific recommendations can be made.

Future Outlook

Nickel aluminide alloys are strong, tough, high-temperature-resistant materials that may find selected applications in wear-critical applications, especially where their corrosion resistance and high-temperature properties may also be a factor. Only very limited friction and wear data exist for these alloys, and more work is needed, including field trials and cost analyses, before widespread commercial tribological applications are likely to appear.

References

1. J. Weir, "Properties of Nickel Aluminide Alloys," Metals and Ceramics Division, Oak Ridge National Laboratory, 1990

2. C.T. Liu, J.O. Stiegler, and F.H. Froes, Ordered Intermetallics, *ASM Handbook,* Vol 2, ASM International, 1990
3. R.K. Williams, R.S. Graves, F.J. Weaver, and D.L. McElroy, Physical Properties of Ni_3Al Containing 24 and 25 Atomic Percent Aluminum, *Proc. Mater. Res. Soc.,* Vol 39, 1985, p 505-512
4. P.J. Blau and C.E. DeVore, Wear-Machinability Relationships in an Ordered Intermetallic Alloy, *Proceedings of the 1991 Wear of Materials Conference,* American Society of Mechanical Engineers, 1991, in press
5. P.A. March, O.F. Karr, and L.L. Curvin, "Laboratory and Field Comparison of Cavitation Erosion Resistance for Base Materials, Weld Overlays, and Coatings," Norris Engineering Laboratory Report No. WR28-1-900-157, Tennessee Valley Authority, 1986
6. M. Johnson, D.E. Mikkola, P.A. March, and R.N. Wright, The Resistance of Nickel and Iron Aluminides to Cavitation Erosion and Abrasive Wear, *Wear,* Vol 140, 1990, p 279-289
7. P.A. March, "Evaluation of Relative Cavitation Erosion Rates for Base Materials, Weld Overlays, and Coatings" Norris Engineering Laboratory Report No. WR28-1-900-157, Tennessee Valley Authority, 1986
8. M. Sahoo, A. Lui, G. Morin, and V.K. Sikka, "Mechanical Properties and Erosion Behavior of Nickel Aluminides," Report MTL 89-68(J), Canada Center for Mineral and Energy Technology (CANMET), 1989
9. B.J. Marquardt and J.J. Wert, Wear Induced Deformation of Ni_3Al, $(Fe, Co)_3V$, and $(Fe, Ni)_3V$, *Proc, Mater. Res. Soc.,* Vol 39, 1985, p 247-254
10. N.R. Bonda and D.A. Rigney, Unlubricated Sliding Wear of Nickel Aluminides at Room Temperature and 400 °C, *Proc. Mater. Res. Soc.,* Vol 133, 1989, p 589-590
11. P.J. Blau and C.E. DeVore, Sliding Behavior of Alumina/Nickel and Alumina/Nickel Aluminide Couples at Room and Elevated Temperatures, *J. Tribology,* Vol 110, 1988, p 646-652
12. P.J. Blau and C.E. DeVore, Temperature Effects on the Break-In of Nickel Aluminide Alloys, *Proceedings of the 1989 Wear of Materials Conference,* American Society of Mechanical Engineers, 1989, p 305-312
13. P.J. Blau and C.E. DeVore, Sliding Friction and Wear Behaviour of Several Nickel Aluminide Alloys under Dry and Lubricated Conditions, *Tribol. Int.,* Oct 1990, p 226-232

Friction and Wear of Titanium Alloys

F.M. Kustas and M.S. Misra, Martin Marietta Astronautics Group

Introduction

TITANIUM ALLOYS offer an attractive combination of high specific mechanical properties, such as modulus/density and strength/density, because of their relatively low density. Titanium alloys also have high toughness and corrosion resistance, making them useful materials for precision mechanism gears, turbine engine components, and biomedical prosthesis devices. However, these alloys have poor fretting fatigue resistance and poor tribological properties. From theoretical calculations, metals with low theoretical tensile and shear strengths exhibit higher coefficients of friction (μ) than higher-strength materials (Ref 1). Within the class of hexagonal close-packed (hcp) structures, titanium has relatively low values for these properties. Consequently, it is expected that titanium would exhibit high frictional values, which has been demonstrated for titanium sliding against itself in air (μ = 60, Ref 1) and vacuum (Fig. 1) (Ref 2).

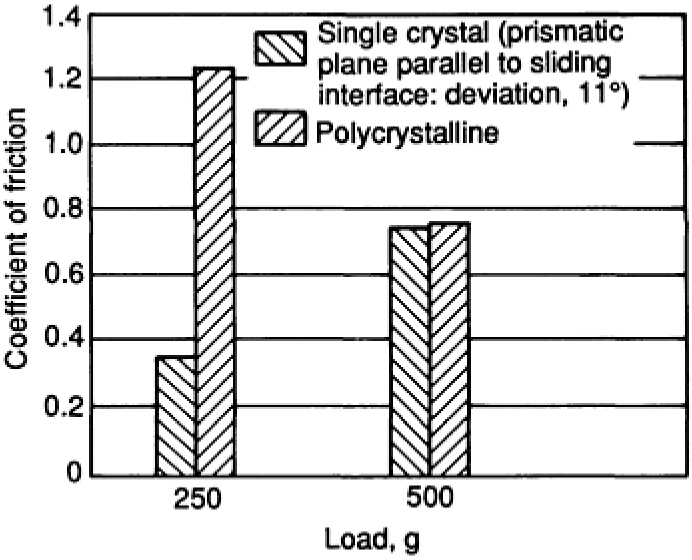


Fig. 1 Coefficient of friction for single-crystal and polycrystalline titanium sliding on polycrystalline titanium in vacuum. Pressure, 1.33 × 10^{-7} Pa (10^{-9} torr); sliding speed, 2.28 cm/s (0.90 in./s). Source: Ref 2

Lower-tensile-strength materials, including titanium, also exhibit greater material transfer to nonmetallic counterfaces than higher-strength metals (Ref 1). The great affinity of titanium for oxygen results in the formation of an oxide surface layer, which is transferred to and adheres to nonmetallic materials, such as polymers, resulting in severe adhesive wear (Ref 2). Addition of a lubricant, such as polyperfluoroalkylether (PFPE) (Ref 3) or kerosene oil (Ref 4), reduces the coefficient of friction and wear damage somewhat, compared with unlubricated conditions Fig. 2) (Ref 3, 4, 5), although reaction of titanium surfaces with these lubricants (generally under high-temperature conditions) can reduce lubricant performance (Ref 6).

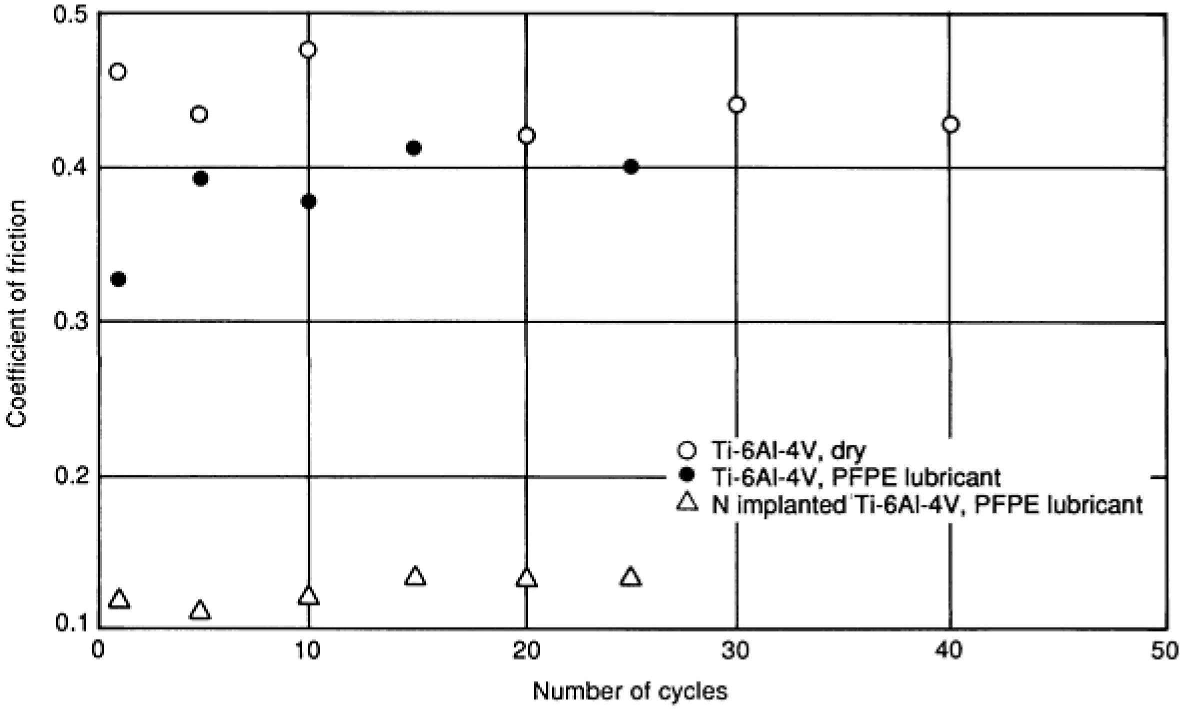


Fig. 2 Reduced coefficients of friction for PFPE-lubricated Ti-6Al-4V compared with unlubricated Ti-6Al-4V in air

environment. Sliding material, WC-Co; sliding speed, 1 mm/s (0.04 in./s); load, 0.5 kgf. Source: Ref 3, 5

Although titanium alloys offer attractive mechanical and physical properties, their surface properties are deficient, thus restricting the use of uncoated titanium alloys to nontribological applications. To realize the full benefit of titanium alloys in friction and wear applications, surface modification treatments are required to effectively increase near-surface strength, thereby reducing the coefficient of friction and lowering the tendency for material transfer and adhesive wear.

Surface Modification Treatments

Many surface treatments have been used to modify the tribological properties of titanium alloys. These treatments can be classified as physical vapor deposition (PVD), including ion implantation (Ref 5, 6, 7, 8, 9, 10, 11, 12, 13), plasma spray (Ref 15, 16), and evaporation (Ref 4); thermochemical conversion treatments, including plasma nitriding (Ref 17, 18), gaseous nitriding (Ref 19), liquid nitriding (Ref 20), ionic nitriding (Ref 21), laser nitriding (Ref 22), ionic carburizing (Ref 21), and laser boriding (Ref 22); plating (Ref 17, 20, 23); and application of solid lubricants by resin bonding/burnishing (Ref 20, 23), direct ion beam deposition to produce diamond-like carbon (DLC) (Ref 24), and sputtering (of molybdenum disulfide and titanium nitride, for example) (Ref 18, 25, 26). These treatments differ dramatically with respect to induced surface temperatures, with nearly ambient conditions for low-dose-rate ion implantation and ion-beam deposition, while temperatures up to 900 °C (1650 °F) are common for plasma nitriding and carburizing treatments. Table 1 summarizes the major advantages and limitations of the various surface modification treatments for titanium alloys.

Table 1 Advantages and limitations of surface modification treatments for titanium alloys

Surface modification technique	Advantages	Limitations
Physical vapor deposition		
Ion implantation	Low-temperature process; effectively net shape; creation of unique near-surface alloys; gradual interface	Line of sight; shallow surface modification zone; high cost
Plasma spray	Rapid deposition rates; thick coatings; inherent porosity can trap lubricant; low cost	Coarse surface finish; need for bond layer, abrupt interfaces; hard-to-control coating composition; line of sight
Evaporation	Large range of available materials; tailorable compositions; high-purity coatings; low cost	Thin coatings; line of sight; abrupt interface (unless ion-beam assisted)
Sputtering	Low-temperature process; potential for layered coatings; tailored coating composition	Thin coatings; line of sight; abrupt interface (unless ion-beam assisted)
Thermochemical conversion		
Nitriding, carburizing, boriding	Gradual interface, conversion of substrate; thick surface modification zones	Potential for hydrogen embrittlement restricts selection of titanium alloys; high process temperatures can distort components and reduced properties; coarse surface finish
Plating	Not line of sight; low cost	Potential for hydrogen embrittlement restricts selection of titanium alloys; abrupt interface

Physical Vapor Deposition

Physical vapor deposition includes ion implantation, sputtering, evaporation, and ion plating surface modification treatments.

Ion implantation is a low-temperature technique for modifying the near-surface region of a material. The modified zone is incorporated within the substrate, leaving no defined interface. The process is a line-of-sight technique and generally offers only shallow penetration depths.

Studies of ion-implanted titanium alloys have concentrated on carbon, nitrogen, and nitrogen-oxygen implantations to produce surfaces enriched in TiC, TiN and Ti_2N, and Ti-O-N, respectively. From pin-on-disk friction wear tests, much greater (that is, two orders of magnitude) wear resistance has been shown (Fig. 3) for nitrogen-implanted Ti-6Al-4V compared with untreated material (Ref 7). Nitrogen or carbon implanted into pure titanium produces nearly equivalent results for breakthrough load to failure of the ion-implanted zoned (defined as applied load for failure of the surface after

fewer than 50 revolutions of the implanted disk) (Fig. 4) (Ref 7). In contrast, nitrogen implantation of higher hardness and higher-strength Ti-6Al-4V is more effective than carbon implantation in increasing breakthrough load (Fig. 4).

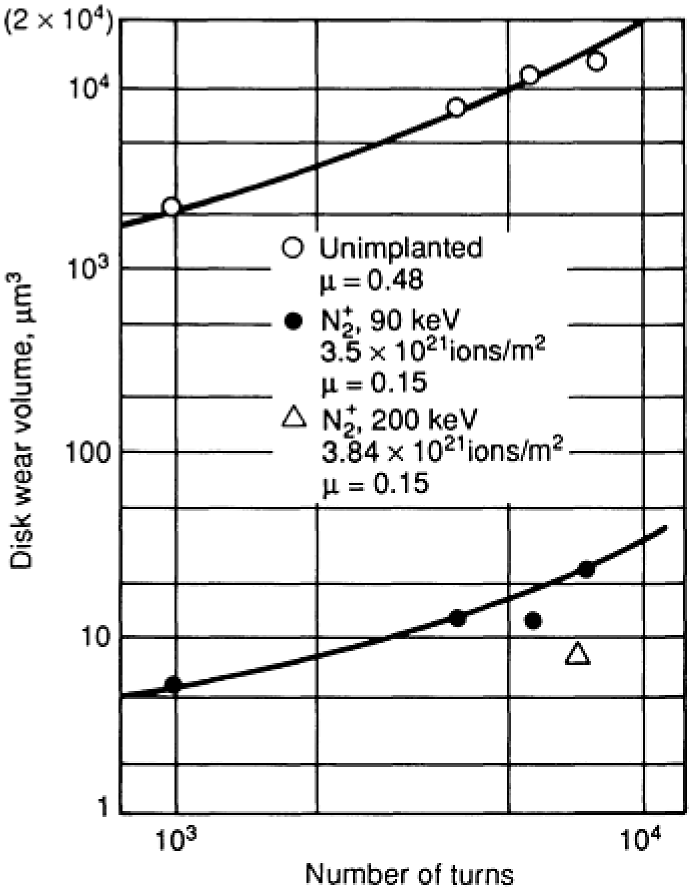


Fig. 3 Wear volume loss as a function of cycles for untreated and nitrogen-implanted Ti-6Al-4V, showing greater than two orders of magnitude reduction in volume loss for nitrogen-implanted material. Pin, 5 mm (0.2 in.) ruby ball; load, 2.61 N (0.266 kgf); velocity, 5.65 mm/s (2.2 in./s). Source: Ref 7

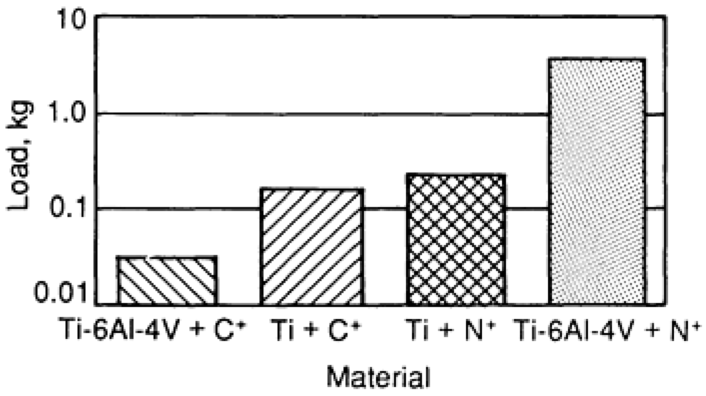


Fig. 4 Breakthrough load for four titaniuim substrate-ion combinations. Pin, 5 mm (0.2 in.) ruby ball; velocity, 56.6 mm/s (2.23 in./s). Source: Ref 7

Additional studies (Ref 8, 9) have shown that (I) an optimum heat treatment of about 400 to 470 °C (750 to 880 °F) after carbon implantation results in greater cycles to breakthrough (an increase by a factor of 30 compared with room-temperature implantation) because of formation of 60 nm diameter TiC particles and (2) a dual-energy implant (to

produce deeper nitrogen penetration) will significantly increase the wear life (by at least one order of magnitude) (Ref 8) of the implanted zone. Carbon implantation increases fretting fatigue life (Fig. 5) by slowing the formation of debris that causes surface damage and subsequent crack initiation (Ref 10).

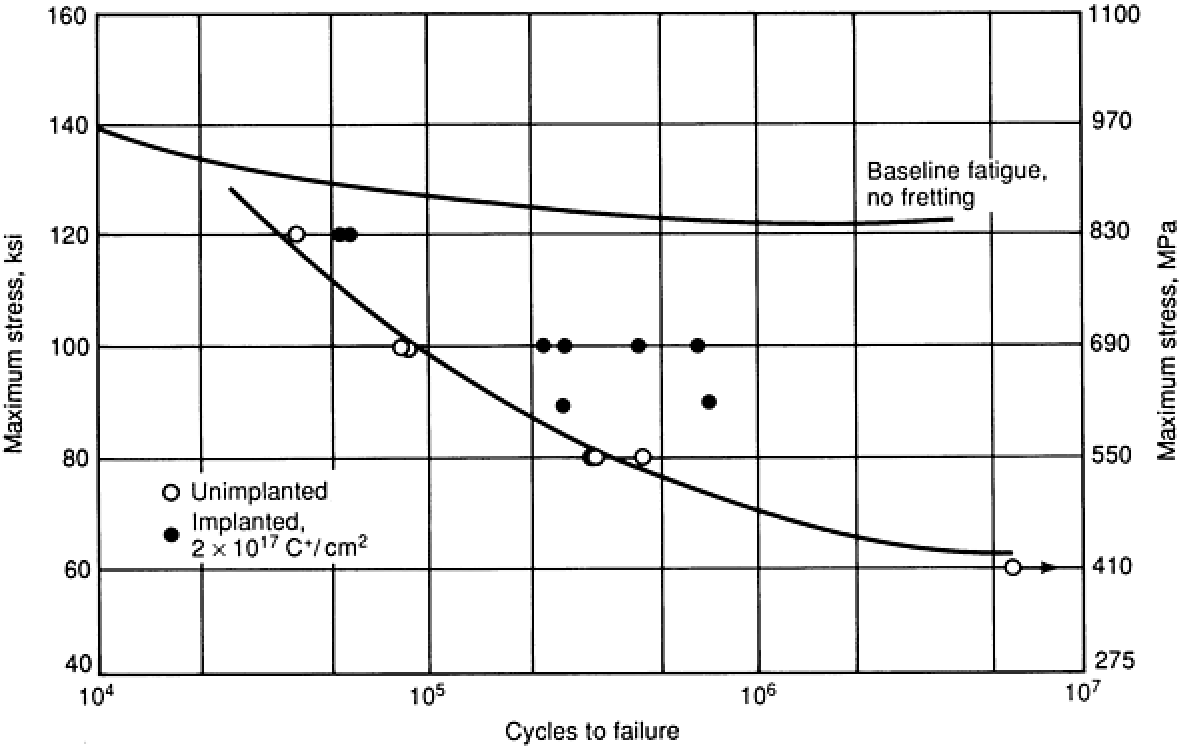


Fig. 5 Cycles to failure versus maximum fatigue stress for carbon-implanted and unimplanted Ti-6Al-4V. Normal stress, 690 MPa (100 ksi); baseline fatigue (no fretting) data, 20.7 MPa (3 ksi). Source: Ref 10, 27

Modification of surface composition enhances the tribological properties of deficient materials. Quantitative analyses of worn surfaces on nitrogen-implanted Ti-6Al-4V have shown that oxygen and carbon are incorporated into the surface layer during either implantation or the wear event. These complex-composition surface layers help produce lower coefficients of friction (Ref 11, 12).

Implantation of Ti-6Al-4V has been performed to improve its corrosive wear resistance when in contact with ultrahigh-molecular-weight polyethylene (UHMWPE) (Ref 13). This combination of materials is used for biomedical prostheses, such as artificial hips (Ref 28), where an integral stem and ball of Ti-6Al-4V is in contact with a UHMWPE acetabular cup. Flakes of TiO_2 formed during wear of titanium in contact with UHMWPE may result in adverse tissue response, including inflammation, infection, loosening of the prosthetic component, and possible carcinogenic reaction. Implantation of Ti-6Al-4V with nitrogen to produce a 20% surface concentration reduced corrosion current in an oxygen-saturated sodium chloride isotonic salt solution by 100 times compared with unimplanted Ti-6Al-4V. The wear surface of the nitrogen-implanted Ti-6Al-4V, in contact with UHMWPE pads, exhibited only a few black stripes after the test--in contrast to the wear surface of unimplanted Ti-6Al-4V, which was completely blackened early in the testing. The authors concluded that the black debris was TiO_2 embedded in the UHMWPE that was transferred to the Ti-6Al-4V surface (Ref 13). In a related study, nitrogen or carbon implantation of Ti-6Al-4V significantly reduced the wear volumes (by five times) of both the pin (Ti-6Al-4V) and the rotating disk (UHMWPE) under a screening environment of Ringer's solution (a lactated 0.9% NaCl solution with the same concentration of Cl^- ions as in body fluids) (Ref 14).

Recently, high-temperature nitrogen implantation of Ti-6Al-4V has been evaluated as a means of increasing the penetration depth of nitrogen into a titanium alloy substrate. It has been demonstrated that the most beneficial tribological properties occur for a 1000 °C (1830 °F) implantation of 10^{18} N_2^+-N^+/cm^2 at 60 keV. This treatment produced the deepest penetration of nitrogen (>750 nm) along with the formation of TiN and Ti_2N (Ref 29). Lowest unlubricated coefficient of friction values (down to about 0.12) were observed for this specimen, with μ decreasing with increasing

applied load until a critical load of 98 N (22 lbf), at which implanted layer failure occurred (Ref 5). Under PFPE lubricated conditions, a threefold reduction in coefficient of friction was observed compared with nonimplanted Ti-6Al-4V for a 600 °C (1110 °F) nitrogen implantation treatment of 10^{17} N_2^+ -N^+/cm^2 at 75 keV (Fig. 2) (Ref 5). The high ion-beam current densities employed during the processing (up to 500 μA/cm^2 versus <50 μA/cm^2 for conventional ion implantation) results in short-time beam exposures of only a few minutes at the elevated temperatures (Ref 29), which would not degrade bulk material properties.

Plasma Spray Coatings. Evaluation of the wear and fatigue performance of spherical bearing components with candidate coatings on Ti-10V-2Fe-3Al showed that plasma spray coatings of Cr_2O_3, Al_2O_3, and WC were the most wear resistant and exhibited the lowest torques (Table 2) (Ref 16). Test results using lubricants contaminated with road dust illustrated the superiority of plasma spray Cr_2O_3 on Ti-10V-2Fe-3Al compared to chromium-plated was 15-5 PH steel where the chromium plating was completely removed during testing. Finally, grease-starved (after initial lubrication) tests showed superior performance for plasma-sprayed Al_2O_3 and WC coatings, which remained intact after 100,000 cycles without addition of grease. The porosity of these coatings served as lubricant reservoirs, providing grease replenishment during the test. As a result, Cr_2O_3-coated titanium components have been incorporated into forward trunnion spherical bearings for aircraft landing gear.

Table 2 Normal lubrication bearing test results for plasma-spray-coated Ti-10V-2Fe-3Al

Bearings were lubricated every 900 cycles with clean grease per MIL-G-23827.

Bearing substrate/coating	Bearing stress		Torque				Cycles	Comments
			At 100 cycles		At end of test			
	MPa	ksi	N · m	lbf · in.	N · m	lbf · in.		
Ti/Ni-B	41	6	220	1950	220	1950	103	Ball locked up in race; coating flaked off
	41	6	. . .	. . .	436	3855	71	Ball locked up in race; coating flaked off
Ti/ion nitride	41	6	220	1950	220	1950	4473	Coating worn off; ball diameter 0.3 mm (0.012 in.) smaller
Ti/Cr_2O_3[(a)]	41	6	259	2295	116.1	1027	10^5	No wear on ball; minimal wear on outer race
	62	9	330	2925	116.1	1027	10^5	No wear on ball; minimal wear on outer race
	90	13	275.5	2438	143.3	1268	10^5	No wear on ball; minimal wear on outer race; outer race shows evidence of heat discoloration
Ti/Cr_2O_3[(b)]	41	6	72.0	637	47.0	416	10^5	No wear on ball; minimal wear on outer race
Ti/Cr	41	6	165.3	1463	84.2	745	10^5	Copper pickup on ball
	62	9	248.3	2197	138.1	1222	10^5	Some surface scoring of chromium plate and copper pickup on ball
Ti/Al_2O_3	41	6	102	903	52.5	465	10^5	Copper pickup on ball
Ti/WC	41	6	144	1270	88.4	782	10^5	Copper pickup on ball
15-5 PH/Cr	62	9	225	2000	110	975	10^5	Copper pickup on ball

(a) Chromium oxide from Source 1.

(b) Chromium oxide from Source 2.

Evaporation. Hard coatings such as TiC, TiN, and Ti_2N deposited on titanium disks by activated reactive evaporation (ARE) have been evaluated against different mating rider materials such as 440C steel, TiC/440C steel, and TiN/440C steel under dry and lubricated (for example, kerosene oil) conditions (Ref 4). In general, the following conclusions can be drawn from the tribological data (Table 3):

- Coefficients of friction and wear volumes (for both rider and disk) are high for both bare and coated 440C steel versus bare titanium
- Coefficients of friction and titanium disk wear volumes are much lower for bare 440C steel versus coated titanium
- Coefficients of friction and wear volumes are lowest for hard coatings on both 440C steel riders and titanium disks

A material combination of TiN/440C rider versus TiC/Ti disk gave the best performance, although the reverse coating combination was not as effective (Table 3).

Table 3 Coefficients of friction and rider and disk wear for various wear couple combinations for uncoated and hard-ceramic-coated titanium

All tests run for 500 m (1640 ft) under a load of 0.4 kg (0.9 lb), unless otherwise noted

Rider	Disk	Coefficients of friction[a]			Rider wear[b]		Disk wear[c]	
		μ_S	μ_D	μ_L	**Dry**	**Lubricated[d]**	**Dry**	**Lubricated[d]**
440C steel, uncoated	Ti, uncoated	0.765	0.45	0.425	High	High	High	High
	TiC/Ti	0.175	0.275	0.275	Medium	High	Low	Low
	TiN + Ti_2N/Ti	0.2	0.35	0.275	High	High	Low	Low
TiC/440C steel	Ti, uncoated	0.65	0.8	0.5	High	. . .	High	. . .
	TiC/Ti	0.275	0.175	0.175	Low	Low	Low	Low
	TiN + Ti_2N/Ti	0.275	0.2	0.1	Low	Low	Low	Low
TiN/440C steel	Ti, uncoated	0.5	0.765	0.4	High	Medium	High	High
	TiC/TiN	0.05	0.025	0.15	Low[e]	Very low	Low	Low

(a) μ_S, static coefficient of friction measured when disk began to slide (velocity: 1.2-1.5 cm/s, or 0.5-0.6 in./s); μ_D, dynamic coefficient of friction measured after about 10 disk revolutions (velocity: 10-15 cm/s, or 4-6 in./s); μ_L, lubricated coefficient of friction measured after about 10 disk revolutions (velocity: 10-15 cm/s, or 4-6 in./s).

(b) Wear test speed was 22-25 cm/s (8.5-10 in./s). Wear volumes are defined as: high, >5 × 10^{-3} mm^3 (3 × 10^{-3} $in.^3$); medium, 10^{-4} to 5 ×10^{-3} mm^3 (0.6 × 10^{-4} to 3 × 10^{-3} $in.^3$); low, 10^{-5} to 10^{-4} mm^3 (0.6 × 10^{-5} to 0.6 × 10^{-4} $in.^3$); very low, <10^{-5} mm^3 (0.6 × 10^{-5} $in.^3$).

(c) Wear test speed was 22-25 cm/s (8.5-10 in./s). Wear volumes are defined as: high, >2.54 × 10^{-3} mm^3 (1.55 × 10^{-3} $in.^3$); medium 2.54 × 10^{-4} to 2.54 × 10^{-3} mm^3 (1.55 × 10^{-4} to 1.55 × 10^{-3} $in.^3$); low, <2.54 × 10^{-4} mm^3 (1.55 × 10^{-4} $in.^3$).

(d) Lubricated with kerosene oil.

(e) Test run for 100 m (328 ft)

Thermochemical Conversion Surface Treatments

Nitriding, Carburizing, and Boriding. Ionic nitriding and carburizing offer the advantage of producing deeper nitride case depths (up to about 6 μm thick) due to thermal diffusion at elevated temperatures (750 to 900 °C, or 1380 to 1650 °F) (Ref 21). These treatments generally are applied to annealed Ti-6Al-4V and pure titanium. An adverse effect of irreversible grain growth was observed to be most prevalent for a 900 °C (1650 °F) treatment (Ref 21). Large-grain microstructures generally have lower mechanical properties than fine-grain material.

Hydrogen-containing compounds, often used in the nitriding process, provide a source of hydrogen that can diffuse into the titanium alloy substrate during the high-temperature nitriding operation. For titanium alloys, the effects of hydrogen contamination on mechanical properties are highly dependent on the specific phase of the titanium alloy microstructure. For example, α-phase titanium alloys are prone to titanium hydride formation (Ref 21), which can embrittle the matrix and thereby reduce mechanical properties. In contrast, titanium alloys with a β-phase microstructure maintain hydrogen in solid solution, which minimizes embrittlement but results in a lower hardening tendency (Ref 21).

Tribological properties of plasma-nitrided titanium alloys are very impressive. Low coefficients of friction (0.05 to 0.08) and low wear rates have been demonstrated for plasma-nitrided Ti-6Al-4V compared with untreated material (Fig. 6) (Ref 18). Fatigue performance of the nitrided material is highly dependent on prior substrate condition. Annealed Ti-6Al-V that was subsequently nitrided showed a marked reduction (up to 21%) in endurance life compared with untreated Ti-6A1-4V; this can be attributed to coarsening of the αgrains and production of a continuous αmatrix (Ref 18). Use of a solution-treated and aged material, however, reduced the fatigue degradation to only 4% for the nitrided material (Ref 18).

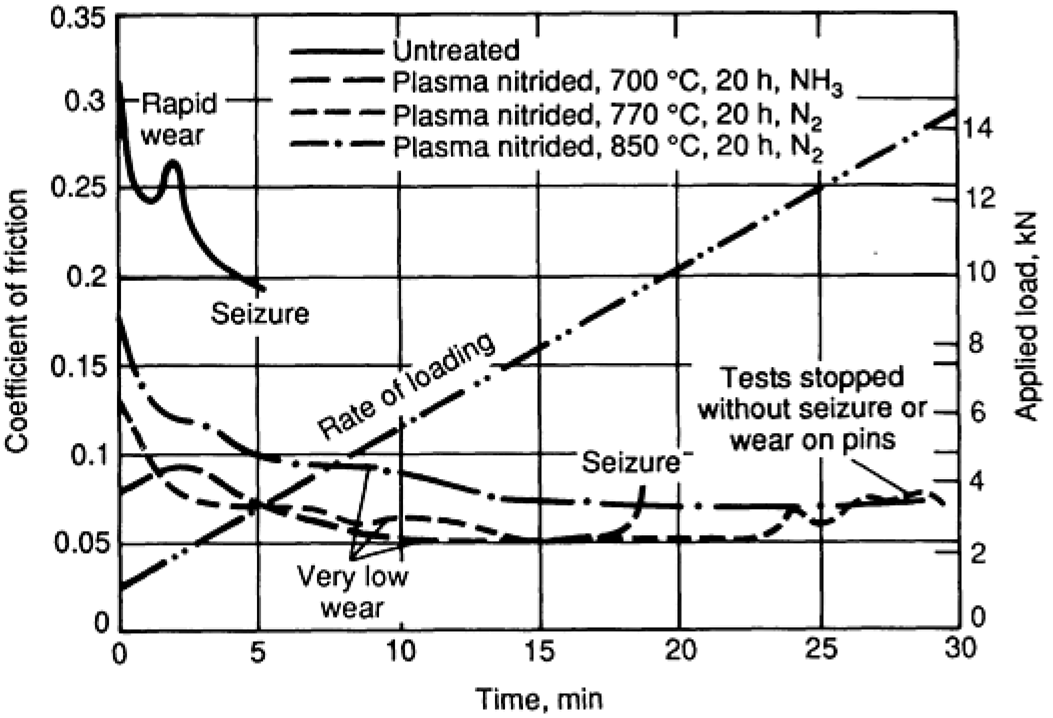


Fig. 6 Wear test results for untreated and plasma-nitrided Ti-6Al-4V. Source: Ref 18

Laser Surface Treatments. Laser surface heating and introduction of hard particles into the surface via reaction with gas (such as nitrogen or methane) or heating of surface deposit (graphite or boron nitride) have been used to produce thick nitride and boride surfaces (Ref 22). Erosion resistance (based on mass loss) of a laser-nitrided Ti-6Al-4V specimen against SiC particles was greater compared with untreated Ti-6Al-4V (Fig. 7). The laser-nitrided layer delays the onset of steady-state erosion that is characteristic for an uncoated Ti-6Al-4V surface. As shown in Fig. 7, the laser-nitrided layer is intact for a lower number of impacts, then suffers fracture and becomes embedded in the substrate as the number of impacts increases. Eventually, a sufficient number of impacts removes the laser-nitrided layer, exposing the uncoated Ti-6Al-4V substrate and resulting in steady-state erosion. In comparison, a TiN layer on a PVD-coated Ti-6Al-4V specimen was removed after only a few impacts.

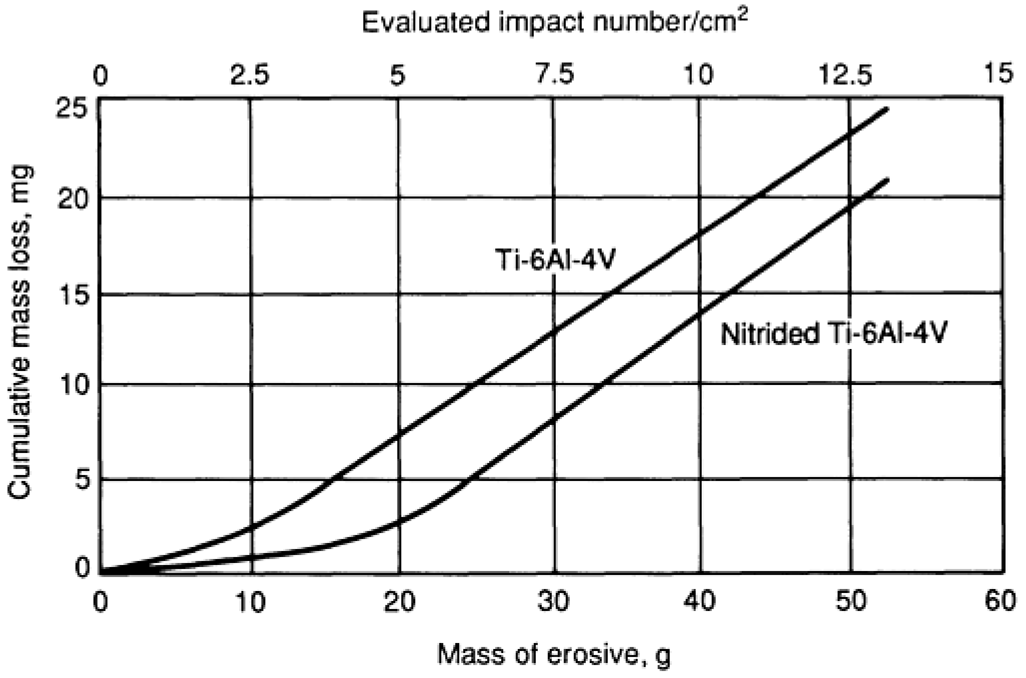


Fig. 7 Cumulative mass loss curves for Ti-6Al-4V and laser-nitrided Ti-6Al-4V. Source: Ref 22

Boride coatings on titanium alloys have shown good resistance to high-velocity particle impacts (Ref 22). Very thick (up to 80 μm), dense coatings have been produced on Ti-6Al-4V by laser boriding. Erosion rates of these thick coatings are reduced, because the material removal mechanism is modified to one of coating cracking rather than erosion of surface material.

Solid Lubrication

Solid lubricants are solid-state materials (generally layered lattice compounds) that are applied to a surface by a burnishing operation (with an organic bonding agent) or deposited by plating or sputtering. Applications for solid-lubricant-coated titanium tribomaterials include components such as compressor blades, uniball bearings, and bearing seal rings for gas turbine engines (Ref 23) and gears for spacecraft mechanism actuators (Ref 17). For the former application, pulse plating deposition of Cr-Mo on Ti-6Al-4V and Ti-8Al-1Mo-1V was performed to produce a solid solution surface layer of Cr-Mo-Ti, which exhibited no degradation of substrate properties, increased creep life by up to 50%, and oxidation protection to 760 °C (1400 °F) (Ref 23). Commercial solid lubricant coatings were subsequently applied to the Cr-Mo-coated/Ti-6Al-4V specimens. Unlubricated wear tests on the solid-lubricant-coated Cr-Mo/Ti-6Al-4V specimens were performed by running them against AISI 1137 steel V-blocks. Test results showed excellent wear resistance for several of the solid lubricant surface treatments (Table 4) (Ref 23). The best performer was PWA 36035 (Fel-Pro C-300: 50% solvent, 47% MoS_2, 3% graphite and other solids). Solvent compatibility tests, however, showed this coating to have poor resistance to jet fuel and oil. Cr-Mo-coated titanium alloy uniballs tested for 520 h in an engine simulator demonstrated no failures, in contrast to a 30% failure rate for standard steel uniballs.

Table 4 Wear tests using Cr-Mo-coated Ti-6Al-4V pins and uncoated steel V-blocks

Solid lubricant	Average time to failure at specified load, min				
	1330 N (300 lbf)	**2220 N (500 lbf)**	**3330 N (750 lbf)**	**4450 N (1000 lbf)**	**Standard deviation**
PWA 474	. . .	. . .	. . .	54	42
PWA 550	. . .	. . .	. . .	63	12
PWA 585	. . .	. . .	. . .	9	0.4
PWA 586	. . .	. . .	. . .	3	[(a)]
PWA 36035	. . .	. . .	. . .	227	29

PWA 36070	1	. . .	. . .	. . .	[(a)]
Lubeco K-350	. . .	. . .	. . .	48	25
Lubeco M-390	. . .	0.2	. . .	. . .	[(a)]
Esnalube 382	. . .	. . .	. . .	6	0.9
Ecoalube 642	. . .	. . .	. . .	74	76

Source: Ref 23

(a) Standard deviations for these tests are misleading, because each set included at least one pin that failed to reach 4450 N (1000 lbf).

Other potential aerospace applications of titanium alloys include drive mechanism components and launch locks (mechanisms used to distribute the load induced by an aerospace vehicle launch through an alternate load path). Replacement of steel gears with titanium alloy gears would reduce subsystem weight and result in reduced launch costs. The effects of various surface treatments on the wear life and efficiency of annealed Ti-6Al-4V gears have been studied by vacuum testing (Ref 17). As shown in Fig. 8, boronized (5 μm layer of metallic borides), Tifran (5 μm layer of complex oxides), and ion-plated aluminum coatings exhibited the best wear resistance and the lowest power losses.

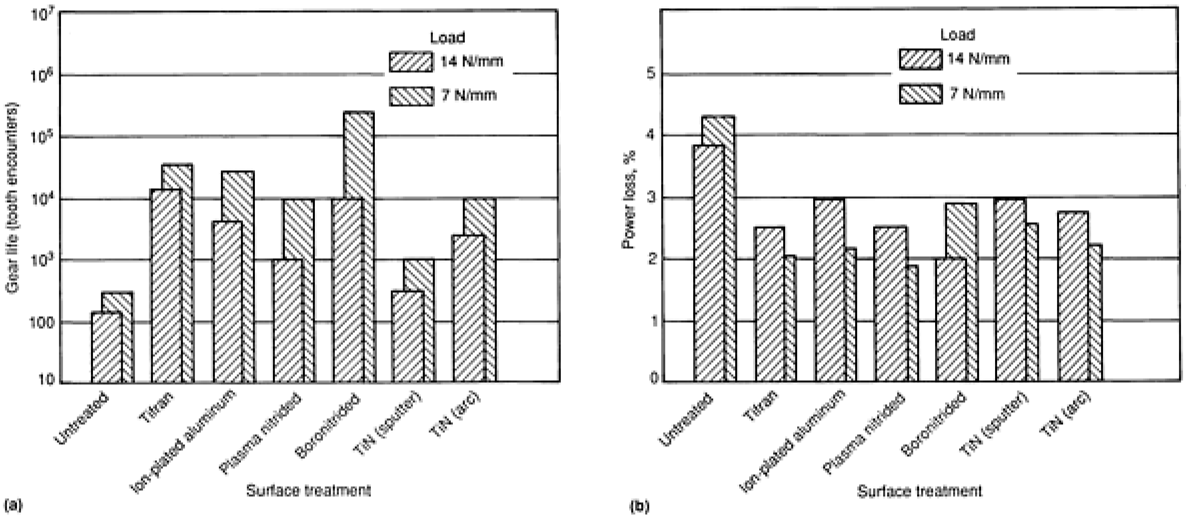


Fig. 8 Performance of surface-treated Ti-6Al-4V gears. (a) Effective life. (b) Effective gear efficiency. Source: Ref 17

The fabrication of solid lubricants such as diamondlike carbon (Ref 24) and molybdenum disulfide (Ref 25, 26) has also been studied. Diamondlike carbon coating exhibits very high hardness and solid lubricant behavior because of its graphite and amorphous carbon contents. To facilitate good DLC adhesion on Ti-6Al-4V, an intermediate bond layer of SiC is required (Ref 24). Unlubricated sliding wear tests in air of DLC/SiC/Ti-6Al-4V against a WC-CO pin showed a gradual reduction in coefficient of friction (down to 0.15) with increased load (Fig. 9) (Ref 24). Diamondlike carbon coating performed very well up to an initial stress level of 1.7 GPa (245 ksi) (52% above the yield strength of the Ti-6Al-4V substrate), at which DLC delamination was observed (Fig. 9).

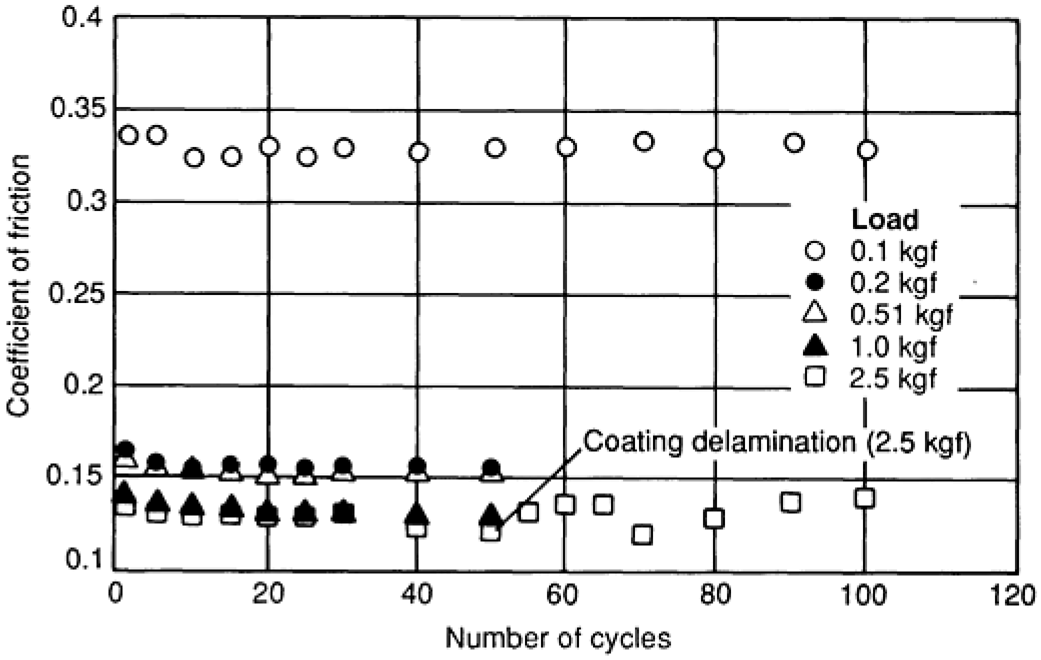


Fig. 9 Coefficient of friction versus wear cycles for 2.5 μm DLC/0.02 μm SiC interlayer coating on Ti-6Al-4V, showing reduction in μ with increased load. Sliding materials, WC-CO; sliding speed, 2 mm/s (0.08 in./s). Source: Ref 24

Coating of titanium alloys with sputtered layered lattice solid lubricants, such as MoS_2, reduces their tendency to gall. Initial studies showed that MoS_2 adhesion to Ti-6Al-4V is very poor, as evidenced by the relatively low vacuum endurance life for MoS_2-coated Ti-6Al-4V loaded against itself (1.7 kgf load, 1.2 m/s sliding speed) of only about 600 revolutions (about 113 m sliding distance) (Ref 25). This is several orders of magnitude lower than for MoS_2-coated steel or ceramic (Si_3N_4) substrates (Fig. 10) (Ref 25).

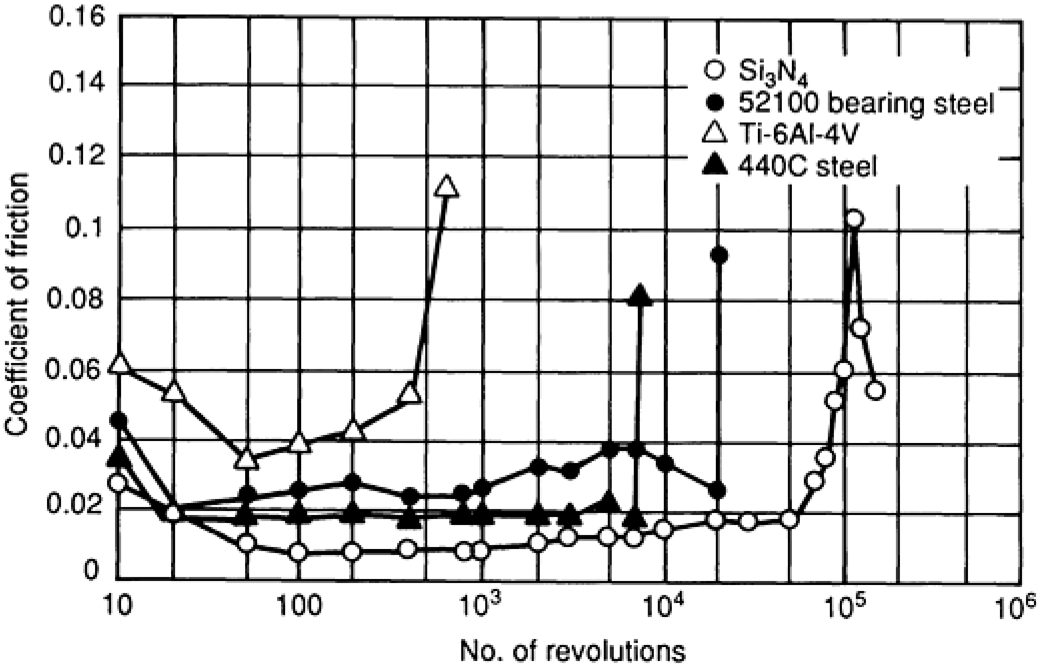


Fig. 10 Endurance lifetimes of MoS_2 on various substrates, showing longest lifetimes for Si_3N_4 and 52100

bearing steel. Sliding material, Ti-6Al-4V; sliding speed, 1.2 m/s (3.9 ft/s); applied load, 17 N/ball (1.7 kgf/ball). Source: Ref 25

Additional research has shown that pretreating the Ti-6Al-4V surface using a nitrogen implantation process produces a ceramic TiN/Ti_2N surface, which significantly increases the load-carrying capacity and high-load endurance lifetime of the sputtered MoS_2 coating on Ti-6Al-4V loaded against steel in an air environment (2 mm/s sliding speed) (Ref 26). Figure 11 shows the relative endurance lifetimes, as a function of applied load, for MoS_2-coated Ti-6Al-4V and MoS_2-coated nitrogen-implanted Ti-6Al-4V specimens. For MoS_2-coated Ti-6Al-4V, a gradual reduction in endurance life from more than 300 reciprocating cycles (5 and 6 kgf loads) to 33 cycles (6.5 kgf load) to 15 cycles (7 kgf load) to nearly instantaneous failure (7.5 kgf load) was observed. In contrast, for the best nitrogen implantation pretreatment of 2×10^{17} $N_2^+ - N^+/cm^2$ at 75 keV (400 °C, or 750 °F, surface temperature), no MoS_2 coating failures were observed for test loads up to the 122.5 N (27.5 lbf) load limit of the test machine. This significant improvement in load-carrying capacity and high-load endurance life has been attributed to improved MoS_2 adhesion to the TiN/Ti_2N-rich Ti-6Al-4V surface through an irradiation-disrupted titanium oxide (Ref 26).

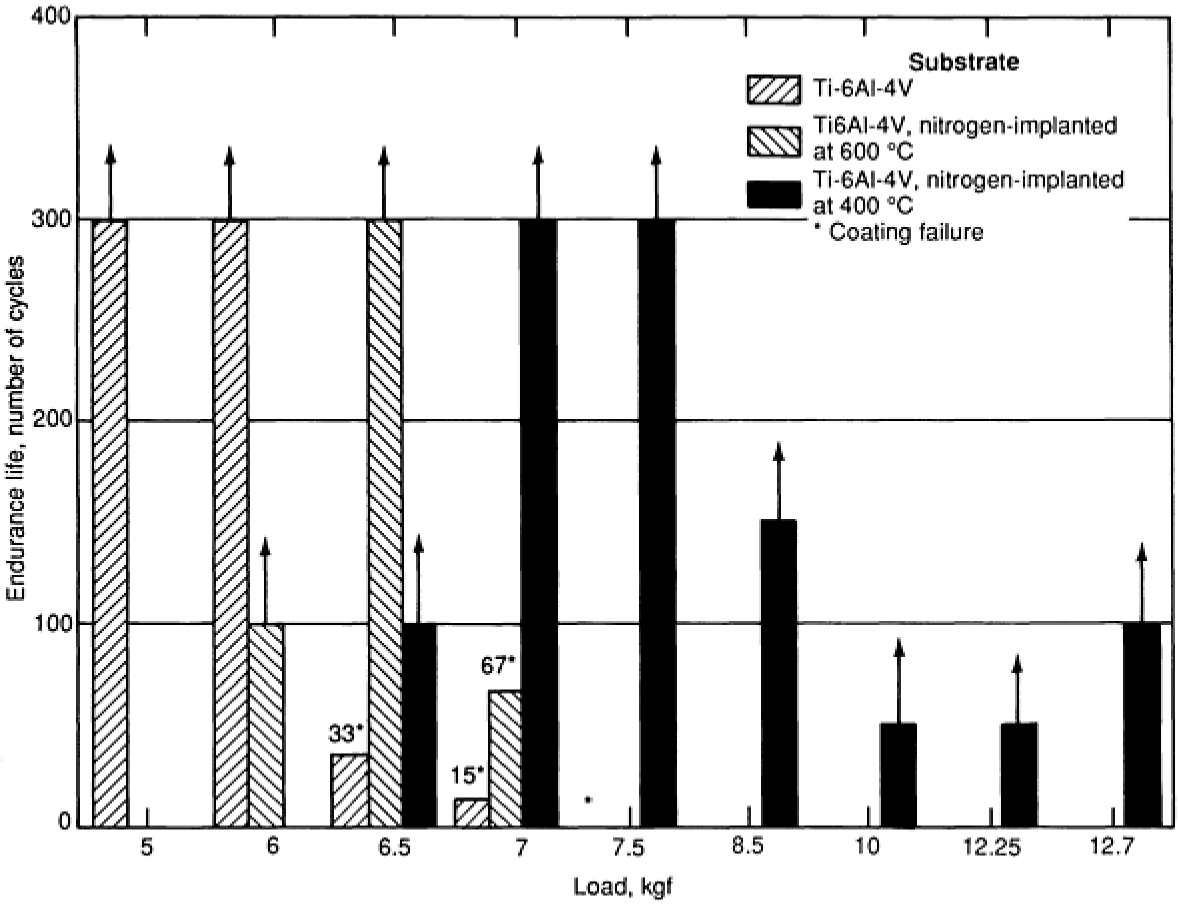


Fig. 11 Endurance lifetimes of MoS_2-coated Ti-6Al-4V specimens, showing significantly increased lifetimes (at high loads) for MoS_2 coating on nitrogen-implanted (processed at 400 °C, or 750 °F) Ti-6Al-4V substrate. Sliding material, bearing steel; sliding speed, 2 mm/s (0.08 in./s). Source: Ref 26

References

1. K. Miyoshi and D.H. Buckley, Correlation of Tensile and Shear Strengths of Metals with Their Friction Properties, *ASLE Trans.*, Vol 27 (No.1), 1982, p 15-23
2. D.H. Buckley, *Surface Effects in Adhesion, Friction, Wear, and Lubrication*, Elsevier Scientific, 1981, p

3. Independent Research and Development Brochure, Project D-81R, Martin Marietta Astronautics Group, 1990
4. T. Jamal, R. Nimmagadda, and R.F. Bunshah, Friction and Adhesive Wear of Titanium Carbide and Titanium Nitride Overlay Coatings, *Thin Solid Films,* Vol 73, 1980, p 245-254
5. F.M Kustas, M.S. Misra, R. Wei, and P.J. Wilbur, High Temperature Nitrogen Implantation of Ti-6Al-4V; Part II: Tribological Properties, *Surf. Coat. Technol.,* 1992, accepted for publication
6. C.E. Snyder, Jr. and R.E. Dolle, Jr., Development of Polyperfluoroalkylethers as High Temperature Lubricants and Hydraulic Fluids, *ASME Trans.,* Vol 19 (No. 3), 1975, p 171-180
7. W.C. Oliver, R. Hutchings, J.B. Pethica, E.L. Paradis, and A.J. Shuskus, Ion Implanted Ti-6Al-4V, *Mater. Res. Soc. Symp. Proc.,* Vol 27, 1984, p 705-710
8. R.G. Vardiman, Wear Improvement in Ti-6Al-4V by Ion Implantation, *Mater. Res. Soc. Symp. Proc.,* Vol 27, 1984, p 699-704
9. R. Martinella, G. Chevallard, and C. Tostello, Wear Behavior and Structural Characterization of a Nitrogen Implanted Ti6Al4V Alloy at Different Temperatures, *Mater. Res. Soc. Symp. Proc.,* Vol 27, 1984, p 711-716
10. R.G. Vardiman, D. Creighton, G. Saliver, A. Effatian, and B.B. Rath, The Effect of Ion Implantation on Fretting Fatigue in Ti-6Al-4V, *Ion Implantation for Materials Processing,* F.A. Smidt, Ed., Noyes Data Corp., 1983, p 165-177
11. B.D. Barton and T.N. Wittberg, "Ion Implantation of Two Titanium Alloys," Report MLM-3603, U.S. Dept. of Energy, Office of Scientific & Technical Information, Aug 1989, p 3
12. F. Pons, J.C. Pivin, and G. Fargas, Inhibition of Tribo-Oxidation Preceding Wear, by Single-Phased TiN_x Films Formed by Ion Implantation into $TiAl_6V_4$, *J. Mater. Res.,* Vol 2 (No. 5), 1987, p 580-587
13. J.M. Williams, G.M. Beardsley, R.A. Buchanan, and R.K. Bacon, Effects of N-Implantation on the Corrosive-Wear Properties of Surgical Ti-6Al-4V Alloy, *Mater. Res. Soc. Symp. Proc.,* Vol 27, 1984, p 735-740
14. P. Sioshansi, R.W. Oliver, and F.D. Mathews, Wear Improvement of Surgical Titanium Alloys by Ion Implantation, *J. Vac. Sci. Technol.,* Vol A3 (No. 6), 1985, p 2670-2674
15. C. Chamont, Y. Honnorat, Y. Berthier, M. Goder, and L. Vicent, "Wear Problems in Small Displacements Encountered in Titanium Alloys Parts in Aircraft Turbomachines," presented at *Sixth World Conference on Titanium* (France), 1988, p 1883-1888
16. R.R. Boyer, Ti/Cr_2O_3 Grease Lubricated Spherical Bearings, *Proceedings of the Technical Program from the 1986 International Conference on Ti Alloys,* Vol I, 1986, p 42-54
17. R.A. Rowntree, Surface-Treated Titanium Alloy Gears for Space Mechanisms, *Proceedings of the Second European Space Mechanisms Symposium* (West Germany), Oct 9-11 1985, p 167-171
18. J. Lanagan, Properties of Plasma Nitrided Titanium Alloys, *Sixth World Conference on Titanium* (France), 1988, p 1957-1962
19. C. Jarbouli, V. Pellerin, D. Treheux, and L. Vincent, Study of the Fatigue Parameters of Titanium Alloys Treated for Wear Resistance, *Sixth World Conference on Titanium* (France), 1988, p 1859-1864
20. M. Thoma, Influence on Fretting Fatigue Behavior of Ti6Al4V by Coatings, *Sixth World Conference on Titanium* (France), 1988, p 1877-1881
21. B. Coll. P. Jacquot, M. Buvron, and J.P. Souchard, "Ionic Nitriding and Ionic Carburizing of Pure Titanium and Its Alloys," presented at ASM International Carburizing Conference, Sept 1989
22. J.P. Massoud and G. Coquerelle, High Power Laser Surface Treatments on Ti-6Al-4V in Order to Improve Its Erosion Resistance, *Sixth World Conference on Titanium* (France), 1988, p 1847-1852
23. P.L. McDaniel, R.E. Fisher, and V.G. Anderson, "Wear Resistant Coatings for Titanium," Report N00019-79-0544, Dept. of the Navy, Naval Air Systems Command, Oct 1980
24. F.M. Kustas, M.S. Misra, R. Wei, and P.J. Wilbur, Diamondlike Carbon Coatings on Ti-6Al-4V, *STLE Tribology Trans.,* 1992, accepted for publication

25. E.W. Roberts and W.B. Price, In-Vacuo, Tribological Properties of "High-Rate" Sputtered MoS_2 Applied to Metal and Ceramic Substrates, *Mater. Res. Soc. Symp. Proc.,* Vol 140, 1989, p 251

26. F.M. Kustas and M.S. Misra, Improved Tribological Performance of MoS_2 on Ti-6Al-4V by Surface Pretreatment, *J. Vac. Sci. Technol.,* 1991, submitted for publication

27. G.L. Goss *et al., Wear,* Vol 24, 1973, p 77

28. J.E. Lemons, Biomaterial Surfaces and Biocompatibility, *Ion Implantation and Plasma Assisted Processes,* ASM International, 1988

29. F.M. Kustas, M.S. Misra, R. Wei, and P.J. Wilbur, High Temperature Nitrogen Implantation of Ti-6Al-4V; Part I: Microstructure Characterization, *Surf. Coat. Technol.,* 1992, accepted for publication

Friction and Wear of Aluminum-Silicon Alloys

Barrie S. Shabel, Douglas A. Granger, and William G. Truckner, Alcoa Technical Center

Introduction

ALUMINUM-SILICON ALLOYS are noted for their unique combination of desirable characteristics, including excellent castability and low density combined with good mechanical properties. Interestingly, there is no mention of wear resistance in the many published tabulations of the attributes of this class of alloys (see, for example, Ref 1 and 2). In a comprehensive review of the wear characteristics of aluminum alloys, Eyre (Ref 3) remarks on the dearth of published information about the wear resistance of aluminum-silicon alloys, while acknowledging their increased application in environments demanding this physical characteristics.

In the late 1950s, die cast aluminum-silicon alloy cylinder blocks were manufactured for the automotive industry to take advantage of the light weight and good thermal conductivity offered by hypoeutectic alloys such as A356 and A380 (Ref 4). However, because these alloys exhibited only modest wear resistance, the early engine blocks (for example, those produced for the American Motors Corporation Rambler) had a cast-in steel cylinder liner. In Europe, automotive engine blocks were also being die cast in aluminum-silicon alloys (such as LM24 and LM26), but these blocks had liners that were shrink-fitted in place. A major boost was given to the use of lightweight aluminum alloys in automobiles when the energy crisis in the early 1970s resulted in the government mandating improved fuel consumption in automobiles. Because fuel consumption can be directly related to vehicle weight (Ref 5), the use of aluminum in place of cast iron in engine components became highly desirable. However, the use of die cast aluminum engine blocks with a cast iron cylinder liner proved to be too costly in production, paving the way for the development of hypereutectic aluminum-silicon alloys with greater wear resistance that could be used without liners.

One such alloy that has been developed specifically for its high wear resistance is the hypereutectic aluminum-silicon alloy A390 (A390.0), which has been used in engine blocks without a liner. An electrochemical surface treatment is used to etch away some of the matrix aluminum so that the eutectic and primary silicon particles provide the bearing surface, resulting in improved wear resistance. Jorstad (Ref 6) has suggested the use of this alloy for other weight-saving applications where good wear resistance is required (for example, brake drums and disk brake rotors).

A powder metallurgy (P/M) production route for the manufacture of cylinder liners from hypereutectic aluminum-silicon alloys has also been explored (Ref 4). Using this method, the size of the primary silicon particles may be optimized and solid graphite can be added to the matrix to serve as a lubricant. The liners are then inserted into the engine block, which is produced from conventional aluminum-base alloys.

While applications for aluminum-silicon alloys are currently centered in the automotive industry (which accounts for greater than 50% of the market), other applications in communications equipment, instrumentation, and small engines appear likely. The hypereutectic alloys are likely to dominate because their wear resistance, combined with low coefficient to thermal expansion (CTE) and fluidity properties, allows thinner wall castings to be manufactured.

Both the Aluminum Association (AA) and the Society of Automotive Engineers (SAE) designations are commonly used to classify aluminum-silicon alloys in the United States.

Metallurgy of Aluminum-Silicon Alloys

Aluminum alloys for wear resistance applications are based on the aluminum-silicon alloy system. This binary system is a simple eutectic alloy system with the eutectic composition at 12.5% Si (Fig. 1). Standard alloys, of course, contain a number of alloying ingredients. Selected commercial alloy compositions are shown in Table 1. Table 2 cross references United States aluminum-silicon compositions with compositions of European and Japanese sources.

Table 1 Nominal compositions of selected commercial alloys recommended for wear applications

Alloy	Composition, wt%					
	Si	**Fe**	**Cu**	**Mg**	**Mn**	**Other**
Hypereutectic alloys						
390.0	16-18	1.3	4-5	0.45-0.65	0.1	. . .
A390.0	16-18	0.5	4-5	0.45-0.65	0.1	. . .
B390.0	16-18	1.3	4-5	0.45-0.65	0.5	1.5 Zn
392.0	18-20	1.5	0.4-0.8	0.8-1.2	0.2-0.6	0.5 Ni, 0.5 Zn, 0.3 Sn
393.0	21-23	1.3	0.7-1.1	0.7-1.3	0.1	2-2.5Ni
Eutectic alloys						
384.0	10.5-12	1.3	3-4.5	0.1	0.5	0.5 Ni, 3 Zn, 0.35 Sn
336.0	11-13	1.2	0.5-1.5	0.7-1.3	0.35	2-3 Ni, 0.35 Zn
339.0	11-13	1.2	1.5-3	0.5-1.5	0.5	0.5-1.5 Ni, 1 Zn
413.0	11-13	2	1	0.1	0.35	0.5 Ni, 0.5 Zn
4032	11-13.5	1	0.5-1.3	0.8-1.3	. . .	0.5-1.3 Ni
Hypoeutectic alloys						
319.0	5.5-6.5	1	3-4	0.1	0.5	0.35 Ni, 1 Zn
356.0	6.5-7.5	0.6	0.25	0.2-0.45	0.35	0.35 Zn
364.0	7.5-9.5	1.5	0.2	0.2-0.4	0.1	0.25-0.5 Cr, 0.15 Ni, 0.15 Sn
380.0	7.5-9.5	2	3-4	0.1	0.5	0.5 Ni, 3 Zn, 0.35 Sn
333.0	8.-10	1	3-4	0.05-0.5	0.5	0.5 Ni, 1 Zn
332.0	8.5-10.5	1.2	2-4	0.5-1.5	0.5	0.5 Ni, 1 Zn
360.0	9-10	2	0.6	0.4-0.6	0.35	0.5 Ni, 0.5 Zn
383.0	9.5-11.5	1-3	2-3	0.1	0.5	0.3 Ni, 3 Zn, 0.15 Sn

Table 2 Cross-reference to equivalent wear-resistant aluminum-silicon alloys

Specific composition limits may vary from United States limits.

United States alloy	Foreign alloy equivalent			
	United Kingdom	**France**	**Japan**	**Germany**
390	LM28	A-S18UNG	. . .	G-AlSi17Cu4
392	. . .	. . .	AC9B	. . .
413	. . .	. . .	. . .	G-AlSi12(Cu)
336	LM13	A-S12UNG	AC8A	. . .
332	LM26	A-S10UG	AC8C	. . .
333	LM2	. . .	AC8B	. . .
360	. . .	A-9G	. . .	G-AlSi10Mg(Cu)
380	LM24	. . .	. . .	G-AlSi8Cu3
356	. . .	A-S7G	AC4C	. . .

Source: Ref 8, 9

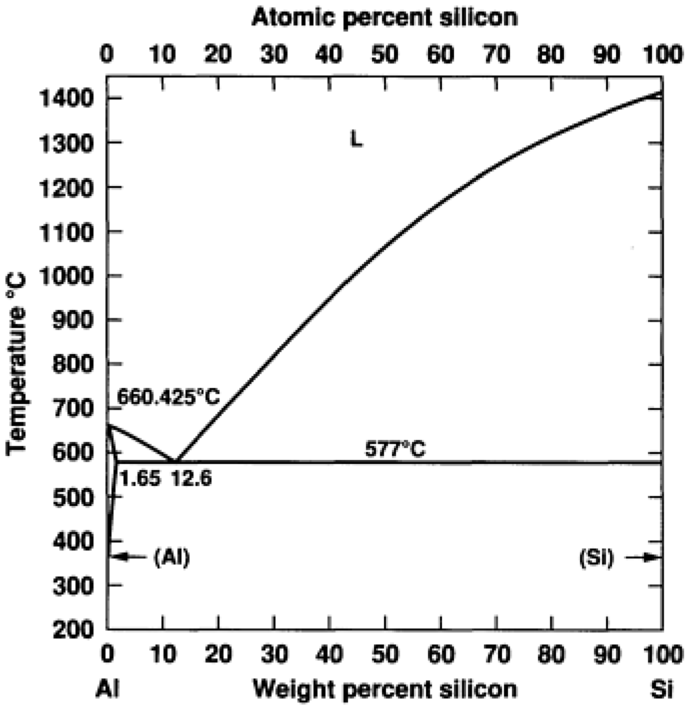


Fig. 1 Aluminum-silicon binary phase diagram. Source: Ref 7

At room temperature, the hypoeutectic alloys consist of the soft, ductile primary aluminum phase and the very hard, brittle silicon phase associated with eutectic reaction. It is this silicon phase that contributes to the very good wear resistance of these alloys. The silicon phase is diamond cubic with a density of ~2.6 g/cm^3 (0.049 lb/in.^{3}) and a Vickers hardness of approximately 10 GPa (1.5 × 10^6 psi). Silicon is essentially insoluble in aluminum (Ref 10). Figure 2 illustrates the typical microstructure of a common hypoeutectic alloy, A357.0. Hypereutectic alloys, the most commonly used wear-resistant alloys, contain coarse, angular, primary silicon particles as well as eutectic silicon. These primary silicon particles impart excellent wear resistance to these alloys. A typical microstructure of an unrefined hypereutectic alloy (A390-type) is shown in Fig. 3.

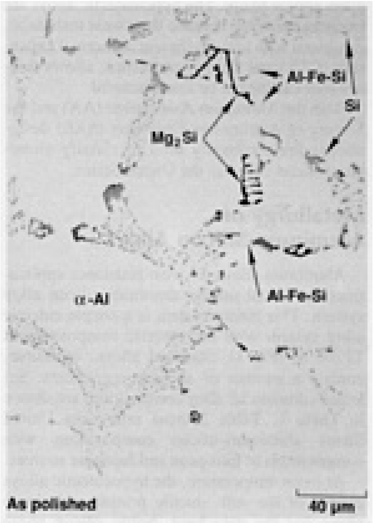


Fig. 2 Typical microstructure of type A357.0 hypoeutectic alloy. Source: Ref 11

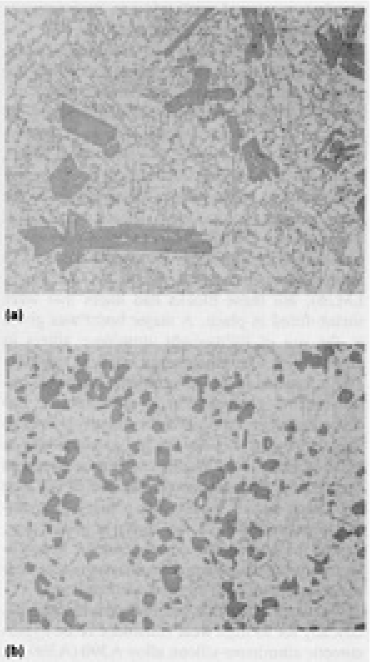


Fig. 3 Microstructure of type A390.0 hypereutectic alloy. (a) Unrefined (Graff-Sargent etch). Dark regions contain coarse primary silicon particles in addition to eutectic silica. (b) Refined (as polished). 120×

Commercial aluminum-silicon alloys (Table 1) generally contain other alloying elements to further enhance or modify the wear resistance or impart additional properties to these alloys.

Iron. The most common alloying element is iron, which can be tolerated up to levels of 1.5 to 2.0% Fe. The presence of iron modifies the silicon phase by introducing several Al-Fe-Si phases. The most common of these are the α and β phases. The α phase has a cubic crystal structure and appears in the microstructure as a "Chinese script" eutectic. The less common β phases generally appear as needles and/or platelets in the structure. Other iron-bearing phases such as Al_6Fe and $FeAl_3$ can also be found in these alloys. Aluminum-silicon alloys intended for die castings typically have higher minimum iron levels to reduce sticking between the mold and the casting.

Magnesium is added to provide strengthening through precipitation of Mg_2Si in the matrix. In an Al-Fe-Si-Mg alloy, the Al-Si-Fe phases will not be affected by the addition of magnesium. However, magnesium can combine with insoluble aluminum-iron phases, resulting in a loss of strengthening potential (Ref 12).

Copper. The most common aluminum wear-resistant alloys also contain copper. Copper additions impart additional strengthening of the matrix through the aging or precipitation-hardening process ($AlCu_2$ or Q phase) or through modification of the hard, brittle Al-Fe-Si phases by substitution in these intermetallic phases. As the strength of these alloys increases through magnesium and copper additions, some sacrifice in ductility and corrosion resistance occurs.

Manganese. Many of the important aluminum-silicon alloys also contain low (<1 wt%), but significant, amounts of manganese. The presence of manganese can reduce the solubility of iron and silicon in aluminum and alter the composition and morphology of the Al-Fe-Si primary constituent phases. For example, manganese additions can favor the formation of constituents such as Al_{12} $(Fe,Mn)_3$ rather than the $Al_9Fe_2Si_2$-type constituents. The manganese-bearing constituents are typically less needlelike or platelike than the manganese-free iron- or (iron/silicon)-bearing primary constituents. Manganese additions also improve elevated-temperature properties of the aluminum-silicon alloys.

Cumulative Effect of Alloying Elements. In summary, aluminum wear-resistant alloys are based on alloys containing the land, brittle silicon phase. Alloying elements such as iron, manganese, and copper increase the volume fraction of the intermetallic silicon-bearing phases, contributing to increased wear resistance compared to binary aluminum-silicon alloys. In addition, magnesium and copper also provide additional strengthening by producing submicroscopic precipitates within the matrix through an age-hardening process.

Properties and Structure

Alloying aluminum with silicon at levels between about 5 and 20% imparts a significant improvement in the casting characteristics relative to other aluminum alloys. As a result, these high-silicon alloys are generally utilized as casting alloys rather than for the manufacture of wrought products. Aluminum-silicon alloys also possess excellent corrosion resistance, machinability, and weldability (Table 3).

Table 3 Relative ratings of aluminum-silicon sand casting and permanent mold casting alloys in terms of castability, corrosion-resistance, machinability, and weldability properties

Aluminum Association number of alloy	Property[a]						
	Resistance to hot cracking[b]	**Pressure tightness**	**Fluidity[c]**	**Shrinkage tendency[d]**	**Resistance to corrosion[e]**	**Machinability[f]**	**Weldability[g]**
Sand casting alloys							
319.0	2	2	2	2	3	3	2
354.0	1	1	1	1	3	3	2
355.0	1	1	1	1	3	3	2
A356.0	1	1	1	1	2	3	2
357.0	1	1	1	1	2	3	2
359.0	1	1	1	1	2	3	1
A390.0	3	3	3	3	2	4	2
A443.0	1	1	1	1	2	4	4
444.0	1	1	1	1	2	4	1
Permanent mold casting alloys							
308.0	2	2	2	2	4	3	3
319.0	2	2	2	2	3	3	2
332.0	1	2	1	2	3	4	2
333.0	1	1	2	2	3	3	3
336.0	1	2	2	3	3	4	2
354.0	1	1	1	1	3	3	2
355.0	1	1	1	2	3	3	2
C355.0	1	1	1	2	3	3	2
356.0	1	1	1	1	2	3	2
A356.0	1	1	1	1	2	3	2
357.0	1	1	1	1	2	3	2
A357.0	1	1	1	1	2	3	2
359.0	1	1	1	1	2	3	1
A390.0	2	2	2	3	2	4	2
443.0	1	1	2	1	2	5	1
A444.0	1	1	1	1	2	3	1

(a) For ratings of characteristics, 1 is the best and 5 is the poorest of the alloys listed. Individual alloys may have different ratings for other casting processes.

(b) Ability of alloy to withstand stresses from contraction while cooling through hot-short or brittle temperature range.

(c) Ability of molten alloy to flow readily in mold fill thin sections.

(d) Decrease in volume accompanying freezing of alloy and measure of amount of compensating feed metal required in form of risers.

(e) Based on resistance of alloy in standard salt spray test.

(f) Composite rating, based on ease of cutting, chip characteristics, quality of finish, and tool life. In the case of heat-treatable alloys, rating is based on T6 temper. Other tempers, particularly the annealed temper, may have lower ratings.

(g) Based on ability of material to be fusion welded with filler rod of same alloy.

Binary hypoeutectic alloys are too soft to have a good machinability rating. However, the machinability of aluminum-silicon alloys is generally very good in terms of surface finish and chip characteristics. Tool life can be short with

conventional carbide tools, particularly in the case of the hypereutectic alloys. With the recent introduction of diamond cutting tools, tool life has been significantly increased, making the machining of the hypereutectic alloys practical.

Corrosion resistance of these alloys is generally considered excellent. Alloys containing increasing amounts of copper have a somewhat lower corrosion resistance than the copper-free alloys as measured in standard salt spray tests.

Because of their high fluidity and good casting characteristics, these alloys are highly weldable with conventional welding techniques. For joining purposes, brazing alloys and filler wire alloys (for example, alloys 4043 and 4047) (Ref 14) are also based on the aluminum-silicon alloy system.

For wear applications, the important physical properties of these alloys include thermal expansion, thermal conductivity, electrical conductivity, and Young's modulus. Data for these properties are available in standard references (Ref 13, 14, 15, 16, 17, 18). Because silicon is generally in precipitate form, the rule of mixtures is applicable when calculating the properties.

Heat Treatment. Depending on the application, thermal treatments can be employed to:

- Increase strength
- Control thermal growth
- Improve ductility

Aluminum-silicon alloys containing copper and magnesium can be heat treated and aged in the same manner as wrought precipitation-hardened alloys. Depending on the strength level required, room-temperature aging (T4 temper) or elevated-temperature aging (<205 °C, or 400 °F) (T6 temper) may be required after heat treatment to obtain the necessary properties. As the strength of the alloy increases from the T4 to T6 temper, reductions in ductility will occur as the strength increases.

In addition, aluminum-silicon and Al-Si-*X* alloys can be given a higher temperature (205 to 260 °C, or 400 to 500 °F) aging treatment from the as-cast condition to improve their strength and thermal stability. This is particularly important for applications where dimensional tolerances are critical (for example, when the alloy is operated at elevated temperatures as a piston component in an engine). Generally, such an aging practice is designated by the T5 temper.

High-temperature (480 to 540 °C, or 900 to 1000 °F) treatments can also be given to aluminum-silicon and Al-Si-*X* alloys to improve their ductility. These thermal treatments modify the angular primary silicon particles to a more rounded shape. This rounded shape reduces the tendency for crack initiation beginning at the sharp edges of the particles. Such treatments are particularly effective on the hypereutectic alloys. Other means of modifying the shape for improved ductility are discussed in the sections "Modification" and "Refinement" in this article.

Principles of Microstructural Control. The three categories of aluminum-silicon alloys are based on the silicon level (Table 1). These alloy categories are hypoeutectic, eutectic, and hypereutectic. The hypoeutectic alloys solidify with α-aluminum as the primary phase followed by aluminum-silicon eutectic. Other solutes (for example, iron, magnesium, and copper) form phases that separate in the freezing range of the alloy in the interdendritic locations (Fig. 2).

Hypereutectic alloys solidify in a similar manner, but in these alloys silicon is the primary phase rather than α-aluminum (Fig. 3a). The eutectic alloys solidify principally with an aluminum-silicon eutectic structure; either aluminum or silicon is present as a primary phase depending on which side of the eutectic composition (12.7% Si) the alloy lies. A brief description of microstructural control is given below; additional information is available in Ref 2 and 12.

Grain Structure. The grain size of the primary aluminum is controlled through the addition of heterogeneous nuclei to the melt in the form of master alloy inoculants such as Al-6Ti or Al-Ti-B (in the latter, the titanium can range from 3 to 5% and the Ti:B ratio from 3:1 to 25:1). Grain sizes vary from $\sim$100 to 500 μm ($\sim$0.004 to 0.020 in.). An example of the effect of grain refinement by an Al-Ti-B refiner is shown in Fig. 4.

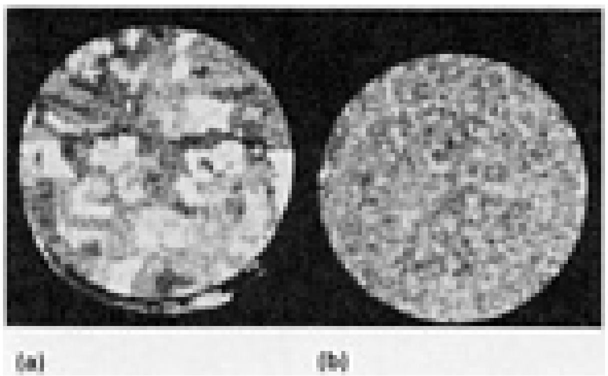


Fig. 4 Effect of grain refinement by the addition of an Al-5Ti-0.2B master alloy to type A356.0. (a) Without titanium addition. (b) With 0.04% Ti addition. Etched with Poulton's reagent. 0.85×

Cell Size. The interdendritic arm spacing (or cell size) is controlled by the cooling rate (Ref 19), which is in turn a function of the casting process and section thickness. The smallest cell size is achieved with thin-wall high-pressure die casting. At the other extreme, thick-wall and castings exhibit the largest dendrite cell size. Casting processes such as low-pressure die casting and permanent mold casting provide intermediate solidification rates and consequently cell sizes that lie between the two extremes. In a similar fashion to the cell size, the constituent phase size is largely controlled by the freezing rate.

Modification. The term modification refers to the change in morphology and spacing of the aluminum-silicon eutectic phase induced by the addition of a chemical agent such as sodium or strontium. There is a change from large divorced silicon particles to a fine coupled aluminum-silicon eutectic structure with an addition of approximately 0.001% Na or 0.005% Sr to the melt. Varying degrees of modification (Fig. 5) are obtained with lower levels of addition. For details of the mechanism and practice of modification, see Granger and Elliott (Ref 12).

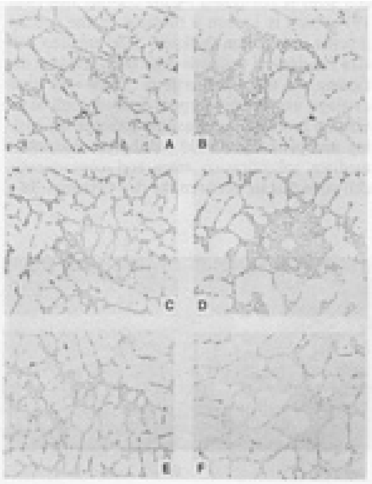


Fig. 5 Variation in microstructure as a function of the degree of modification. The modification level increases from A to F; thus microstructure F is highly modified. Source: Ref 2, 12

Antimony is also used to modify (more accurately, refine) the eutectic structure in hypoeutectic and eutectic alloys, particularly in Europe and Japan. Like sodium and strontium, it increases the fluidity of the alloys and improves mechanical properties. Furthermore, it is permanent, allowing melts to be more effectively degassed, which, in turn, provides sounder castings. The great disadvantage of antimony is that it poisons (or negates) the effect of sodium and strontium, and it also creates a problem in recycling. An additional serious drawback is the potential for the formation of stibine gas, which is highly toxic. Unlike sodium and strontium, which can be used to effectively modify eutectic

structures over a wide range of freezing rates, antimony provides eutectic refinement only at the relatively high rates experienced in die castings and some thin-wall permanent mold castings.

Refinement. In hypereutectic alloys, the primary phase is silicon. In order to provide the desired small well-dispersed silicon particles, phosphorus is added to the level of about 0.1% P through the addition of a master alloy such as Cu-10P. The phosphorus combines with aluminum to form aluminum phosphide, AlP, which provides effective nuclei for the silicon phase much the same as TiB_2-type particles are effective nuclei for α-aluminum (Fig. 3b). However, phosphorus also negates the effectiveness of sodium and strontium. It does so by combining with them to form phosphides, which do not modify the eutectic structure. Similarly, sodium and strontium reduce the effectiveness of phosphorus additions by refining the primary silicon phase.

Gas Porosity. Hydrogen porosity can be controlled by maintaining gas levels at $\leq$0.10 cm^3/100 g. This is not readily accomplished, particularly when modification of the melt is being sought with the addition of sodium or strontium. However, gas fluxing methods are available (Ref 20) that provide the means of reducing hydrogen levels to the desired range.

Also deleterious to casting soundness is the presence of nonmetallic inclusions that act as nuclei for gas pores. Various molten metal filtration systems are available for inclusion removal (Ref 20).

Sludge. A problem experienced with aluminum-silicon alloys is the formation of hard intermetallic phases of the Al(FeM)Si-type, which settle out under gravity from the melt (Fig. 6). The conditions that favor the formation of these phases are low holding temperatures (which are often employed in the die-casting industry); a quiescent melt; and relatively high levels of iron, manganese, and chromium. The relative tendency to form sludge in the holding furnace is given by a segregation factor (SF):

$$SF = (\%Fe) + 2\,(\%Mn) + 3\,(\%Cr) \quad \textbf{(Eq 1)}$$

The relationship among the segregation factor, holding temperature, and sludging tendency is given for alloy AA 339 and several other aluminum-silicon alloys in Ref 21.

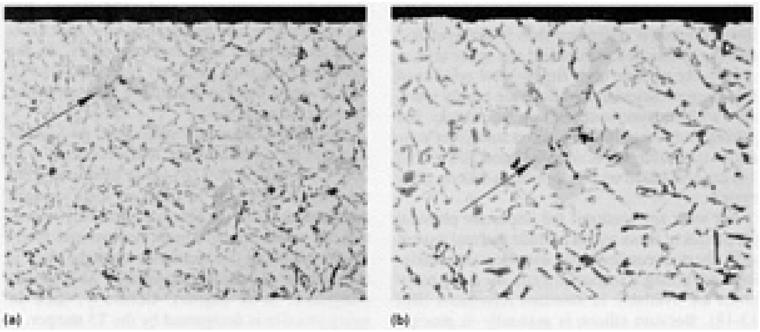


Fig. 6 Coarse intermetallic Al_{12} (Fe,Mn,Cr)$_3$ Si_2 phase constituent generated by entrapped sludge in alloy 339. (a) 130×. (b) 265×

Wear Behavior

The two major types of wear relevant to industrial applications of aluminum-silicon alloys are "abrasive" and "sliding" wear. These have also been identified by Eyre (Ref 3) as the most common types of wear. Wear mechanisms, though, can be thought of as involving more specific descriptions of local processes occurring in the metal and countersurface of the wear system during the wear process. Wear mechanisms are discussed in detail in this Volume in the Sections "Wear by Particles or Fluids," "Wear by Rolling, Sliding, or Impact," and "Chemically Assisted Wear." The purpose of this discussion is to focus on the interaction between microstructure and wear mechanisms. This is important for aluminum-

silicon alloys because of the variety of microstructures that can be achieved as the alloys are processed for particular applications. The relative effects of silicon particles, matrix hardness, and intermetallic constituents on the wear resistance of aluminum-silicon alloys are summarized below.

Silicon Particles. Under relatively light load conditions, which are normally associated with low (<10^{-11} m^3/m) losses, wear resistance is not a strong function of silicon content (Ref 22, 23, 24, 25). In general, however, silicon additions to aluminum will increase the wear resistance. The principal mechanism appears to be the influence of the hard silicon particles, which lead to higher overall levels of hardness. The fact that the hard silicon particles are surrounded by a softer and relatively tough matrix improves the overall toughness of the material and can contribute to wear resistance by favoring more plastic behavior. The eutectic and hypereutectic silicon alloys, with increased volume fractions of hard primary silicon particles relative to the hypoeutectic alloys, might be expected to have the best wear resistance of the aluminum-silicon alloys. Andrews *et al.* (Ref 26, 27), for example, found that increasing the silicon content in hypereutectic alloys reduced wear. However, binary alloy data (Ref 22, 28) indicate that the hypereutectic alloys are not necessarily the most wear resistant. Clarke and Sarkar (Ref 28) found that there was a relative minimum in wear for binary aluminum-silicon alloys at about the eutectic level, as did Jasim *et al.* (Ref 22), especially at applied pressures <100 kPa (<15 psi). Clarke and Sarkar attribute the effects of silicon in part to its effect on metal transfer mechanisms between the pin and countersurface (Ref 29). There is also evidence for increased wear resistance with refinement of the silicon particle morphology (Ref 30, 31).

It is clear, therefore, that microstructure-based explanations are needed to account for the variation in wear rates with silicon content. Moreover, there is a need to account for the reduction in strength that occurs with increased silicon content (Ref 32, 33). The complex effects of composition on wear behavior suggest that wear resistance depends on other material properties (for example, fracture toughness) (Ref 34). Thus lower fracture toughness at higher levels of silicon could lead to higher wear rates if larger pieces of debris are created during the wear process. Variations in toughness and strength with composition might also account for the apparent ability of the near-eutectic compositions to have a greater load-bearing capability at a given wear rate than either higher or lower silicon levels.

Matrix Hardness. Increased matrix hardness is typically achieved through the heat treatment response produced by copper and magnesium additions. Most commercial applications of aluminum-silicon alloys, in fact, depend on the increased strength achieved by heat treatment. The improved wear resistance of precipitation-strengthened material compared to solid solution strengthened material under low wear conditions was also noted by Soderberg *et al.* (Ref 35) using aluminum alloy 6061, which is strengthened primarily by Mg_2 Si precipitates. This is also the strengthening mechanism in the heat-treatable magnesium-bearing aluminum-silicon alloys. Although heat treatment has a beneficial effect (Ref 26, 27, 32), variations in matrix hardness may be less important than the effects of silicon content (Ref 27).

Intermetallic Constituents. In addition, there are important "other" hard phases present in commercial aluminum-silicon alloys that provide enhanced wear resistance. These constituents (for example, Al-Fe-Si, Al-Fe-Mn, Al-Ni, Al-Ni-Fe, Al-Cu-Mg) have varying degrees of hardness (Ref 36, 37, 38). Despite the apparent scatter, these constituents are all much harder than the aluminum matrix. Some examples of the hardness values of these intermetallic compounds are shown in Table 4.

Table 4 Typical hardness values of selected intermetallic constituents of aluminum-silicon alloys

Phase	Hardness		Ref
	MPa	**kfg/mm²**	
$CuAl_2$	3900	400	36, 37
	3800-7600	390-780	36
$FeAl_3$	7200	730	36, 37
	6400-9400	650-960	37
	5160-7110	526-725	36
	3500	360	38
$NiAl_3$	6000-7600	610-770	36
	7100	720	37
	4500	460	38
Ni_2Al_3	9800-11,000	1000-1120	36, 37
Si	7000-14,200	715-1450	36
	11,880	1211	37
Mg_2Si	4480	457	37

Al_2CuMg	3700-3900	380-400	37
Al_9FeNi	8400-9680	860-987	37
$Al_{12}Fe_3Si$	10,760	1097	37

Typical room-temperature hardness values for the aluminum alloy matrices would be <1000 MPa (<100 kgf/mm^2). The hardness values of the intermetallics decrease with increasing temperature (Ref 36), albeit at slower rates than the matrix hardness.

The addition of "hard" phases in the form of particles or fibers to reduce wear is also utilized to create metal-matrix composites (MMC) materials (Ref 39). These materials utilize hard intermetallic, cermet, or ceramic phases to provide the high hardness material for wear resistance. Hornbogen (Ref 40) and Zum Gahr (Ref 41) have described in quantitative terms how the contribution of these hard phases to wear resistance can be modeled in terms of their volume fraction and morphology. This composite approach has been effectively used to develop new piston materials (see the section "Metal-Matrix Composites" in this article).

Finally, the use of "softer" constituents (for example, graphite) should also be noted as an active area for development of wear-resistant aluminum-silicon MMC materials (Ref 32, 39, 42). In these materials, ranking may depend on whether volumetric wear rates (in units of m^3/m) or seizure resistance is being considered. The presence of the softer phase may lead to greater volumetric wear in some cases but greater resistance to seizure (higher load at seizure) in other cases.

To summarize, the results of wear studies using aluminum-silicon alloys illustrate a variety of mechanisms. The effect of variations in silicon particle morphology is often not clear cut, although heat treatment is beneficial to the sliding wear resistance. Therefore, selection of an optimum microstructure is often difficult in practical situations where several wear types or mechanisms could occur. In general, either eutectic or hypereutectic alloys offer the greatest wear resistance under a wide range of wear conditions. Selection may then hinge on the dependence of in-service performance on other alloy characteristics or cost. Overall, the aluminum-silicon alloy system provides a good basis for developing lightweight, strong, wear-resistant materials. Examples of these applications will be discussed in the following sections.

Aluminum-Silicon Alloy Applications

Aluminum-silicon alloys are used in a variety of automotive, aerospace, and consumer product applications.

Automotive Components

Table 5 lists typical automotive components made from aluminum-silicon casting alloys (Ref 43). The eutectic or nearly eutectic alloys (for example, 332, 336, and 339) (Ref 44), are perhaps the most widely used. Equivalent versions of these alloys are used for similar applications by European and Japanese automakers (Ref 45, 46, 47).

Table 5 Automotive engine applications of aluminum-silicon alloys

SAE alloy	Type of casting[a]	Typical application
319.0	S	General purpose alloy
332.0	PM	Compressor pistons
333.0	PM	General purpose
336.0	PM	Piston alloy (low expansion)
339.0	PM	Piston alloy
355.0	S, PM	Pump bodies, cylinder heads
390.0	D	Cylinder blocks, transmission pump and air compressor housings, small engine crankcase, air conditioner pistons
A390.0	S, PM	Cylinder blocks, transmission pump and air compressor housing, small engine crankcase, air conditioner pistons

Source: Ref 43

(a) S, sand cast; D, die cast; PM, permanent mold.

Pistons. Typical applications for aluminum-silicon alloys in the French automotive industry are shown in Table 6 (Ref 45). In addition to being cast, the A-S12UN (eutectic) alloy can also be forged (Ref 48). Similarly, AA 4032, somewhat

similar in composition to 336, is also widely used as a piston alloy (for example, for high-performance forged pistons). Hypereutectic alloys are also used for cast pistons, especially in diesel engines (Ref 45, 49). The potential benefit from composites that combine the strength reinforcement of ceramics with an aluminum-silicon alloy matrix has also been evaluated (Ref 45, 47, 48).

Table 6 Selected aluminum-silicon alloy applications in automobiles produced in France according to engine type and specific automobile manufacturer

Engine type	Manufacturer			
	Citroen	**Peugot**	**Renault**	**Talbot**
Gas	A-S12UN	A-S10UN(F)[(a)]	A-S12UN	A-S10.5UN
	. . .	A-S12UN(A)[(b)]	. . .	A-S11UN
	. . .	. . .	. . .	A-S12UN
Diesel	A-S18UN	A-S12UN	A-S18UN	. . .
	. . .	A-S13UN	. . .	. . .

(a) F, cast iron liner.

(b) A, aluminum block.

Engine Blocks and Cylinder Liners. The evolution of lightweight power plants has depended not only on lightweight pistons but also on the availability of wear-resistant cylinder liners and engine blocks. Hypereutectic liners were described by Mazodier (Ref 50) and El Haik (Ref 51). It was also known that hypereutectic aluminum-silicon alloys had excellent properties for engine blocks (Ref 52, 53, 54). This led to the development and application, in both the United States and Europe, of the A390 (A-S17U4) alloys for die cast engines (Ref 55, 56, 57, 58, 59, 60). An important aspect of the A390 success is the use of a "system" (Ref 60) that includes the engine alloy, the piston materials (electroplated cast F332[AA 332.0] alloy), and the cylinder bore finishing process. Fine honing to a 0.075 to 0.15 μm (3 to 6 μin.) surface finish, followed by controlled etching/polishing to leave silicon particles standing slightly above the alloy surface, was deemed necessary for optimum wear resistance.

Efforts to simplify the 390-type technology by finding a more wear-resistant alloy for the cylinder or reducing the difficulties of finishing the bore have led to substitute alloys. One approach has been the use of a lower silicon alloy containing more nickel and manganese (for example, the Australian-3HA alloy, with a nominal composition of Al-13.5Si-0.5Fe-0.45Mn-0.5Mg-2Ni) (Ref 61).

Continuing interest in the use of more highly wear-resistant materials in other engine-related parts has led to recent applications such as roller-type valve rocker arms (Ref 62) and valve lifters for the Toyota Lexus (Ref 63, 64). The rocker arm alloy used in the Mazda 929 is a nominal Al-10Si-2.7Cu-0.8Mg-0.45Mn alloy somewhat similar to the AA 383 alloy. The valve lifter, on the other hand, is a strontium-modified Al-Si-Cu alloy designated 4T12 (composition, Al-10.5Si-4.5Cu-0.6Mg-0.2Mn).

Typical examples from a more detailed compilation of aluminum alloys used in wear-resistant applications in U.S. autos are shown in Table 7.

Table 7 Wear-resistant aluminum-silicon alloys used in automotive piston components produced for United States automotive manufacturers in 1978 to 1985 model years

Application	Manufacturer	Model/make	Model year(s)
Internal combustion engine components			
Pistons	American Motors	All	1978-85
	Ford	Mercury	1978-81, 83-85
	General Motors	Buick	1978-81, 84-85
		Others	1978-81, 83-85
Brake system components			
Wheel cylinder pistons	American Motors	All	1983-85
	Chrysler	All	1983-85

	Ford	All	1978-85
	General Motors	All (except for Cadillac) Cadillac	1984-85
		Cadillac	1985
Master cylinder pistons	American Motors	All	1984-85
	Chrysler	All	1984-85
	Ford	All	1983-85
	Ford	Mercury	1978-81, 84-85
	General Motors	All	1984-85
Transmission components			
Intermediate band servo pistons	Ford	Some	1983-85
Rear band servo pistons	Chrysler	Some	1984-85
	Ford	Some	1983-85

Source: Ref 65

Bearing Alloy Components. Aluminum alloys have been utilized for bearing applications for many years. The many uses range from diesel and internal combustion engines to a variety of tooling applications (for example, presses, lathes, and milling machines) (Ref 66). Important cast bearing alloys were based on aluminum-silicon or Al-Sn-Cu alloys, whereas wrought bearing alloys have included the 8*xxx* types (for example, AA 8081 and AA 8020) (Ref 66, 67). Compositions of various aluminum bearing alloys are listed in Table 8.

Table 8 Nominal compositions of standard aluminum-silicon alloys used in bearing applications

Alloy		Composition, wt%					
Aluminum Association designation	**SAE designation**	**Si**	**Sn**	**Cu**	**Fe**	**Ni**	**Cd**
8.50	770	0.7	5.5-7	0.7-1.3	0.7	0.7-1.3	. . .
8280	780	1-2	5.5-7	0.7-1.3	0.7	0.2-0.7	. . .
851	. . .	2-3	5.5-7	0.7-1.3	0.7	0.3-0.7	. . .
852	. . .	0.4	5.5-7	1.7-2.3	0.7	0.9-1.5	. . .
. . .	781	3.5-4.5	. . .	0.05-0.15	0.35	. . .	0.8-1.4
8081	. . .	0.7	18-22	0.7-1.3	0.7	. . .	. . .
. . .	782	0.3	. . .	0.7-1.3	0.3	0.7-1.3	2.7-3.5
. . .	783	0.5	17.5-22.5	0.7-1.3	0.5	0.1	. . .

Source: Ref 66, 67, 68

Improved strength and fatigue performance, as well as some increased wear resistance, has been achieved with silicon additions. Thus, alloys SAE 780 and SAE 781 have become widely used for automotive applications such as main and connecting rod bearings (Ref 68, 69). The higher silicon alloy, 781, is also used in bushings and thrust bearings. Its improved wear performance has been attributed to the increased silicon content of the wear surface (Ref 70). These aluminum-silicon alloys are readily used with steel backing in high-load situations.

Advanced Aluminum Bearing Alloys. The nominal compositions of improved bearing alloys with silicon additions are listed in Table 9.

Table 9 Nominal compositions of advanced aluminum-silicon alloys used in bearing applications

Alloy[(a)]	Composition, wt%						Ref
	Si	**Sn**	**Cu**	**Mg**	**Pb**	**Other**	
A	11	. . .	1	. . .	. . .	. . .	6
B	3	10	0.4	. . .	1.8	0.3 Cr	8
C	12	. . .	1	1.5	. . .	3 C, 1 Ni	9
D	. . .	. . .	4.5	. . .	. . .	3 C	. . .
E	11	. . .	. . .	. . .	20	1.4 In	10
F	2.5	12	1	. . .	. . .	0.25 Mn	11
G	6	. . .	1.2	0.5	1	4 Zn	12

H	4	0.5	0.1	0.1	6	0.3 Mn	13

(a) Arbitrary designations

Although a soft phase (for example, tin) is normally considered desirable for avoiding seizure, the compatibility of an Al-11Si-1Cu alloy was better than that of the traditional aluminum-tin alloy (SAE 783) (Ref 71). The silicon-copper alloy also had much better fatigue resistance. The improved properties resulted in applications such as diesel crankshaft and connecting rod bearings. Nevertheless, in line with the concerns expressed by Davies (Ref 72), the harder silicon-containing alloy was more sensitive to misalignment-induced seizure.

The addition of silicon to aluminum-tin alloys containing lower tin levels than that of the 783 alloy provided a compromise between the conformability of the soft-phase material and the benefits of the harder silicon phase for improved wear and fatigue resistance (Ref 73). This alloy could apparently be used without the common lead alloy overlays employed for seizure resistance. The importance of fatigue resistance was also emphasized in the improved Al-Sn-Si alloys reported by Ogita *et al.* (Ref 74). As shown in Table 9, these alloys are somewhat similar to those of Fukuoka *et al.* (Ref 73).

As another alternative to the lead- or tin-containing alloys, a graphite-containing Al-12Si alloy has been successfully evaluated for bearings (Ref 75). The use of a modified lead plus indium addition to an Al-11Si-Pb bearing alloy has also been reported (Ref 76).

Japanese concerns with the pollution and toxicity aspects of cadmium-containing alloys have led to improved aluminum-zinc bearing alloys (Ref 77). These have also been improved with silicon additions. The additional matrix wear between the silicon particles is believed to create lubricant reservoirs that enhance seizure resistance. However, overlays (lead-tin alloys) are still required for best conformability.

Finally, the combined effect of silicon and refinement of the silicon- and lead-bearing phases by rapid solidification processing has resulted in an improved Al-6Pb-4Si bearing alloy (Ref 78). This alloy has grown in usage recently and is projected to be used in 78% of the cars built in the United States during 1991.

Because of the increasing understanding of the balance among wear, fatigue, and seizure resistance of bearing materials, silicon alloys have been used to develop new and improved bearing materials. Further improvements will undoubtedly be necessary as engine operating conditions evolve toward higher temperatures and operating speeds.

Consumer Electronics Components

The growth of this market, which encompasses video cassette recorders (VCRs), video tape recorders (VTRs), digital audio tape (DAT) applications, and other devices (for example, personal computers), has created numerous opportunities for the use of lightweight, relatively corrosion-resistant and wear-resistant aluminum-silicon alloys. VTR cylinders are specifically cited by various Japanese authors (Ref 79, 80, 81) because the eutectic-type silicon alloys have low coefficients of friction against the tape.

Aerospace Components

A nonautomotive engine application of aluminum-silicon alloys is the use of the 390-type alloys in an aircraft engine (the Thunder engine) (Ref 82).

Breakthroughs in Aluminum-Silicon Wear-Resistant Materials

Metal-Matrix Composites. Composite pistons were recognized early as a potentially viable application of MMC technology. While some composite approaches, especially for the severe operating conditions of diesel pistons, recommended the addition of specific metallic inserts to achieve improved performance (Ref 83), the bulk of development efforts have gone into the incorporation of ceramic fibers.

The use of ceramic fiber aluminum-silicon MMC materials for pistons is described in a variety of publications (Ref 47, 48, 84, 85, 86, 87, 88). There are clear benefits to strength at elevated temperatures and reduction of the thermal expansion coefficient. These materials appear to be especially applicable in critical areas such as the top piston rings and

top land (a high-temperature area). The castability of the aluminum-silicon alloys is a favorable factor in their use as matrices, particularly because squeeze casting is one of the preferred fabrication routes for composite pistons.

The property improvements at elevated temperatures have encouraged ongoing development of the MMC technology for automotive engine applications, including engine blocks. The ability of the MMC approach to allow selective strengthening of the cylinder region was taken advantage of by Honda engineers (Ref 89), who utilized composite reinforcement of alloy ADC12 (a Japanese alloy similar to 383)in the manufacture of a die cast engine block.

Powder Metallurgy. The combined benefits of high silicon content and refined silicon particle size on wear resistance are strong driving forces behind the use of P/M techniques for making aluminum-silicon alloy parts. The P/M approach has been of special benefit to the hypereutectic silicon alloys. One example of this is the use of P/M A-S17U4 alloy to make cylinder liners (Ref 90, 91). The properties of these alloys exceed those of standard alloys (Ref 92). In addition, high levels of additional elements can be utilized to obtain good strength and wear-resistant properties at elevated temperatures (Ref 93, 94, 95).

The refined microstructures available from the P/M fabrication of hypereutectic alloys have a beneficial effect on fatigue characteristics as well. This attribute has been utilized in the production of rotors and vanes for rotary automotive air conditioners (Ref 96). For this application, P/M alloys with high levels of iron or nickel are blended with P/M 2024-type alloys to create alloys containing 17 to 20% Si and 5 to 8% Fe or Ni.

Spray casting, as exemplified by Osprey processing, has been shown to offer benefits similar to powder metallurgy (Ref 97, 98). This has the potential for even greater cost savings, which is an important factor if the aluminum industry is to compete successfully in the automotive market. In a comparison of the structure in an Al-20Si alloy, Kahl and Leupp (Ref 97) showed that there was practically no difference in silicon particle size between the P/M and Osprey processing routes. Furthermore, they claimed the the fine and uniform silicon particle size resulted in improved wear behavior compared with that of conventionally produced material, although details of their test procedure are not know. Earlier work at Delft University (Ref 99) compared the rate of mass loss of an Al-20Si-3Cu-1.3Mg alloy rubbing against cast iron at pressure level of 5.5 MPa (800 psi). In this test, the Osprey material showed better resistance to wear than either the ingot metallurgy (I/M) or P/M samples. The reason for the better performance of the Osprey product compared with the P/M material was not clear, but it was hypothesized that the slightly larger silicon particles of the Osprey product helped reduce the fretting wear.

Coatings/Surface Treatments. Other approaches to the wear (adhesion) problems of aluminum pistons moving in aluminum cylinders have taken the path of coating or surface treatment of the piston rings and cylinder bore to minimize wear problems (Ref 100). The selective fiber strengthening noted above is related to this problem also. One widely used treatment is the so-called Nikasil treatment (Ref 101), an electrochemical treatment utilizing a dispersion of silicon carbide particles, preferably <4 μm (<160 μin.) in size, in a nickel matrix.

Efforts to take advantage of both the P/M and surface treatment approach are exemplified by the emerging surface treatment technologies of surface alloying via ion implantation (Ref 102, 103, 104, 105), thermal spraying (Ref 106, 107, 108), and surface treating or alloying using laser treatments (Ref 109, 110, 111). Ion implantation is recognized for its ability to impart wear-resistant surfaces, but principal applications have been to protect tool surfaces in critical processing operations. Laser hardening can be achieved through laser cladding to produce a chemically different surface or through the effective heat treatment (or remelting and rapid solidification) brought about by laser heating. The work by Blank *et al.* (Ref 111) using 7 and 12% Si alloys, however, indicated that surface alloying (for example, with iron or iron plus vanadium) was more effective for increased wear than surface heat treatment effects alone. In any case, the ability to tailor surface properties to a technological need will enable engineers to obtain further enhancement of the wear resistance of the aluminum-silicon base alloys without sacrificing their other advantages.

References

1. J.P. Lyle and D.A. Granger, *Ullman's Encyclopedia of Industrial Chemistry,* Vol 1A, VCH Publishers, 1985, p 481-528
2. E.L. Rooy, Aluminum and Aluminum Alloys, *Casting,* Vol 15, 9th ed., *Metals Handbook,* ASM International, 1988, p 743-770
3. T.S. Eyre, Wear Resistance of Metals, *Treatise on Materials Science and Technology,* Vol 13, Douglas Scott, Ed., *Wear,* Academic Press, 1979, p 363-442

4. A. Kaye and A. Street, *Die Casting Metallurgy,* Butterworths Scientific, London, 1982, p 32
5. S. Rabinowitz, The Expanding Role of Aluminum in the Automotive Industry, *Die Cast. Eng.,* Sept-Oct 1, 1977
6. J.L. Jorstad, "Aluminum Lightweight Castings--Some Cost Saving Ideas," Paper No. 770322, SAE Automotive Engineering Congress, Society of Automotive Engineers, 1977
7. J.L. Murray and A.J. McAlister, The Aluminum-Silicon System, *Bull. Alloy Phase Diagrams,* Vol 5 (No. 1), 1984, p 74-84, 89-90
8. Registration Record of Aluminum Association, Alloys Designations and Chemical Composition Limits for Aluminum Alloys in the Form of Castings and Ingot, 1987-09-01, Aluminum Association
9. Registration Record of Aluminum Association, Limits for Wrought Aluminum and Wrought Aluminum Alloys, 1987-04-01, Aluminum Association
10. L.F. Mondolfo,*Aluminum Alloys: Structure and Properties,* Butterworths, 1976
11. D.A. Granger, R.R. Sawtell, and M.M. Kersker, Effect of Beryllium on the Properties of A357.0 Castings,*Trans. AFS,* 1984, p 579-586
12. D.A. Granger and R. Elliott, Aluminum-Silicon Alloys, *Casting,* Vol 15, 9th ed., *Metals Handbook,* ASM International, 1988, p 159-168
13. J.E. Hatch, *Aluminum: Properties and Physical Metallurgy,* American Society for Metals, 1984, p 344-345
14. *Source Book on Selection and Fabrication of Aluminum Alloys,*American Society for Metals, 1978
15. A.L. Kearney and E.L. Aluminum Foundry Products, *Properties and Selection: Nonferrous Alloys and Special-purpose Materials,* Vol 2, *ASM Handbook,* ASM International, 1990, p 123-151
16. A.L. Kearney, properties of Cast Aluminum Alloys, *Properties and Selection: Nonferrous Alloys and Special-Purpose Materials,* Vol 2, *ASM Handbook,* ASM International, 1990, p 152-177
17. J.E. Hatch, Chapters 8 and 9, *Aluminum: Properties and Physical Metallurgy,* American Society for Metals, 1984, p 152-177
18. General Information--Chemical Compositions, Mechanical and Physical Properties of SAE Aluminum Casting Alloys--SAE J452, June 1983, *SAE Handbook,* Society of Automotive Engineers, 1989, p 10.5-10.13
19. R.E. Spear and G.R. Gardner, Dendrite Cell Size, *Trans. AFS,* Vol 71, 1963, p 209-215
20. D.V. Neff, Non-Ferrous Molten Metal Processes, *Casting,* Vol 15, 9th ed., *Metals Handbook,* ASM International, 1988
21. J. Gobrecht, Gravity-Segregation of Iron, Manganese and Chromium in an Aluminum-Silicon Casting Alloy, Part I, *Giesserei,* Vol 62 (No. 10), 1975, p 263-266; and D.A. Granger, Investigation of a Star-Like Intermetallic Phase Occurring in Alloy 339, *Trans. AFS,* Vol 91, 1991, p 379-383
22. K. Mohammed Jasim and E.S. Dwarakadasa, Wear in Al-Si Alloys under Dry Sliding Conditions, *Wear,* Vol 119, 1987, p 119-130
23. B.N. Pramila Bai and S.K. Biswas, Effect of Load on Dry Sliding Wear of Aluminum-Silicon Alloys, *ASLE Trans.,* Vol 29 (No. 1), 1986, p 116-120
24. R. Shivanath, P.K. Sengupta, and T.S. Eyre, Wear of Aluminum-Silicon Alloys, *Wear of Materials--1977,* International Conference on Wear of Materials, W. Glaeser, K. Ludema, and S. Rhee, Ed., American Society of Mechanical Engineers, p 120-126
25. R. Antoniou and D.W. Borland, Mild Wear of Al-Si Binary Alloys during Unlubricated Sliding, *Mater. Sci. Eng.,* Vol 93, 1987, p 57-72
26. J.B. Andrews, M.V. Seneviratne, K.P. Zier, and T.R. Jett, The Influence of Silicon Content on the Wear Characteristics of Hypereutectic Al-Si Alloys, *Wear of Metals--1985,* K. Ludema, Ed., American Society of Mechanical Engineers, p 180-184
27. J.B. Andrews, M.V. Seneviratne, and T.R. Jett, The Influence of Si and Other Alloying Additions on the Wear Characteristics of Hypereutectic Al-Si Alloys, *4th European Tribology Congress (EUROTRIB 85),* Vol III, 1985
28. J. Clarke and A.D. Sarkar, Wear Characterization of As-Cast Binary Aluminum-Silicon Alloys, *Wear,* Vol

54, 1979, p 7-16

29. J, Clarke and A.D. Sarkar, The Role of Transfer and Back Transfer of Metals in the Wear of Binary Al-Si Alloys, *Wear,* Vol 82, 1982, p 179-185
30. T.K.A. Jaleel, N. Raman, S.K. Biswas, and K.K. Murthy, Effect of Structural Modification and Load on the Wear of a Hypereutectic Aluminum-Silicon Alloy, *Aluminum,* Vol 60 (No. 12), 1984, p 932-933
31. Y. Odani, K. Akechi, and N. Kuroishi, High Strength and High Wear Resistance Al-Si P/M Alloys, *Modern Developments in Powder Metallurgy,* Vol 16, 1984, p 493-505
32. S. Das, S.V. Prasad, and T.R. Ramachandran, Microstructure and Wear of Cast (Al-Si Alloy) Graphite Composites, *Wear,* Vol 133, 1989, p 173-187
33. B.N. Pramila Bai, S.K. Biswas, and N.N. Kumtekar, Scanning Electron Microscopy Study of Worn Al-Si Alloy Surfaces, *Wear,* Vol 87, 1983, p 237-249
34. K.-H. Zum Gahr, Microstructure and Wear of Materials, *Tribology Series,* Vol 10, Elsevier, 1987
35. S. Soderberg, U. Bryggman, and A. Canales, Influence of Precipitation and Solution Strengthening on Abrasive Wear Resistance, *Wear of Metals--1985,* K. Ludema, Ed., American Society of Mechanical Engineers, p 645-653
36. E.R. Petty, Hot Hardness and Other Properties of Some Binary Intermetallic Compounds of Al, *J. Inst. Met.,* Vol 89, 1960-1961, p 343-349
37. A.A. Ivanko and G.V. Samsonov, *Handbook of Hardness Data,* Translation from Russian, National Bureau of Standards-National Science Foundation, 1968
38. H.E.N. Stone, The Oxidation Resistance, Hardness and Constitution of Metallic Aluminides, *J. Mater. Sci.,* Vol 10, 1975, p 923-934
39. S.V. Prasad and P.K. Rohatgi, Tribological Properties of Al Alloy Particle Composites, *J. Met.,* Nov 1987, p 22-26
40. E. Hornbogen, Description of Wear of Materials with Isotropic and Anisotropic Microstructure, *Wear of Metals--1985,* K. Ludema, Ed., American Society of Mechanical Engineers, p 477-484
41. K.-H. Zum Gahr, Abrasive Wear of Two Phase Metallic Material with a Coarse Microstructure, *Wear of Metal--1985,* K. Ludema, Ed., American Society of Mechanical Engineers, p 45-57
42. B.N. Pramila Bai, E.S. Dwarakadasa, and S.K. Biswas, Scanning Electron Microscopy Studies of Wear in LM13 and LM13-Graphite Particulate Composite, *Wear,* Vol 76, 1982, p 221-220
43. General Information--Chemical Compositions, Mechanical and Physical Properties of SAE Aluminum Casting Alloys, *SAE Handbook,* J452, June 1983, Society of Automotive Engineers
44. R. Wilson, Piston Improvement Relies on Material Science, *Diesel Prog. North Am.,* Vol 52 (No. 2), Feb 1986, p 18-20
45. J. Ayel, The Influence of Metallurgical and Geometric Parameters on the Wear of Recent Engines, *Ing. Automob.,* Sept-Oct 1980, p 82-98
46. M. Coulondre and J. Travaille, New Trends in the Design and Manufacture of Pistons and Cylinders for Automobile Engines, *Ing. Automob.,* Sept-Oct 1980, p 68-72
47. A.R. Baker and S. Gazzard, Developments in Materials for Pistons, *Mater. Des.,* Vol 9 (No. 1), Jan-Feb 1988, p 28-33
48. J.C. Avezou and A. Parker, Evaluation of Materials for Pistons, Segments and Valve Seatings for Petrol and Diesel Engines of Automobiles, *Ing. Automob.,* Oct 1983, p 73-80
49. J.P. Whitacre, "Hypereutectic Aluminum--A Piston Material for Modern High Specific Output Gasoline Engines," TP 871944, Society of Automotive Engineers, 1987
50. M. Mazodier, Aluminum Engine Blocks with Thin Hypersilicon Alloy Liners, *Ing. Automob.,* No. 10, Oct 1977, p 586-590
51. R. El Haik, The Possibilities of Using Al Cylinder Blocks--Hypersiliconized Cylinder Liner Tests, *Ing. Automob.,* No. 10, Oct 1977, p 569-577
52. J.L. Jorstad, "Development of the Hypereutectic Al-Si Die Casting Alloy Used in the Vega Engine Block," TP No. 105, 6th International Society of Die Casting Engineers Congress (Cleveland, OH), Nov 1970

53. J.L. Jorstad, The Hypereutectic Aluminum-Silicon Alloy Used to Cast the Vega Engine Block, *Mod. Cast.,* Vol 60 (No. 4), Oct 1971, p 59-64
54. J. Ayel, Frictional Behavior of the Aluminum Hypersilicon Alloys Used as Engine Liners, *Rev. Alum.,* Apr 1978, p 186-191
55. H.H. Hofmann and K. Schellmann, "Aluminum 390 Alloy Engine Blocks: Design and Manufacturing," TP 830007, Society of Automotive Engineers, 1983
56. W. Dahm and R.G. Putter, "Light Alloy Engines--Experience Gained from Research and Development," TP 830005, Society of Automotive Engineers, 1983
57. G. Renninger, D. Abendroth, and M. Bolien, "Castings Engine Blocks in GK Al Si 17 Cu 4 Mg," TP 830003, Society of Automotive Engineers, 1983
58. E.G. Jacobsen, "General Motors 390 Aluminum Alloy 60 degree V6 Engine," TP 830006, Society of Automotive Engineers, 1983
59. H.M. Ward, III, "Engineering Control for High Volume 390 Die Casting," TP 830009, Society of Automotive Engineers, 1983
60. J.L. Jorstad, "Reynolds 390 Engine Technology," TP 830010, Society of Automotive Engineers, 1983
61. J. Barresi, R. Giannini, and B.J. Clarsen, "Cold Scuff Evaluation of 3HA--The Engine Alloy Developed in Australia," TP 871203, Society of Automotive Engineers, 1987
62. S. Komatsu, S. Yahata, K. Fukui, and T. Tokoshima, "Development of Aluminum Alloy Roller Type Valve Rocker Arm by Casting and Forging Technique," TP 890554, Society of Automotive Engineers, 1989
63. S. Ezaki, M. Masuda, H. Fujita, S. Hayashi, Y. Terashima, and K. Motosugi, "Aluminum Valve Lifter for Toyota New V-8 Engine," TP 900450, Society of Automotive Engineers, 1990
64. H. Baker, The Road Ahead for Metals in Autos, *Adv. Mater, Proc.,* May 1990, p 27-34
65. Aluminum Association Guide to Aluminum Parts 1978-1984, Reynolds Aluminum 1985 Aluminum Automotive Applications
66. Chapter 20, *Aluminum--Vol II: Design and Application,* K.R. Van Horn, Ed., American Society for Metals, 1967
67. Chapters 8 and 9, *Aluminum--Vol I: Properties, Physical Metallurgy and Phase Diagrams,* K.R. Van Horn, Ed., American Society for Metals, 1967
68. "Bearing and Bushing Alloys," J460e, *SAE Handbook,* Society of Automotive Engineers
69. "Bearing and Bushing Alloys," J459c (Information Report), *SAE Handbook,* Society of Automotive Engineers
70. M. Gallerneaut and C. Fulton, "Score Resistance of SAE 781 Automotive Journal Bearing Alloy-A Metallographic Analysis," TP 860062, Society of Automotive Engineers, 1986
71. G.C. Pratt, "Aluminum Based Crankshaft Bearings for the High Speed Diesel Engine," TP810199, Society of Automotive Engineers, 1981
72. F.A. Davis, Plain Bearing Wear in Internal Combustion Engines, *Automot. Eng. (UK),* Vol 6 (No. 4), 1981, p 31-33, 37-38
73. T. Fukuoka, H. Kato, and S. Kamiya, Aluminum Alloy Bearings Containing Hard Particles Fitted for Use with Nodular Iron Shaft, *Studies of Engine Bearings and Lubrication,* SP-539, Society of Automotive Engineers, 1983, p 33-44 (TP 830308)
74. Y. Ogita, K. Niwa, and Y. Kondo, "New Engine Type Fatigue Phenomenon of Al-Based Engine Bearings," TP 890556, Society of Automotive Engineers, 1989
75. B.P. Krishnan, N. Raman, K. Narayanaswamy, and P.K. Rohatgi, Performance of Aluminum Alloy Graphite Bearings in a Diesel Engine, *Tribol. Int.,* Vol 16 (No. 5), 1983, p 239-244
76. D.P. Howe and J.D. Williams, An Al-Si-Pb-In Plain Bearing Material, *Proceedings Fourth Conference of Irish Durability and Fracture Committee* (IDFC4), 1986, p 183-191
77. M. Sakamoto, Y. Ogita, Y. Sato, and T. Tanaka, "Development of New Al-Zn-Si Bearings for Heavy Load Applications in Uprated Engines," TP 900124, Society of Automotive Engineers, 1990

78. *Automot. Eng.,* Vol 97 (No. 12), 1989, p 21-25
79. S. Kitaoka, H. Kamio, and S. Kobayashi, Material Properties of Al-Si Alloys for Forging and Their Applications, translation from *Alutopia (Jpn.),* Vol 17 (No. 10), 1987, p 9-18
80. S. Kitaoka, C. Fujikura, and A. Kamio, Aluminum-Silicon Alloys, translation of *J. Jpn. Inst. Light Metals,* Vol 38 (No. 7), 1988, p 426-446
81. O. Takezoe and Y. Yasuda, Technical trends of Highly Wear Resistant Aluminum Alloy, *Kobe Res. Dev.,* Vol 38 (No. 4), Oct 1988, p 23-26
82. R.P. MacCoon and R.P. Ernst, "The Thunder Aluminum 390 Alloy Aircraft Engine," TP 830008, Society of Automotive Engineers, 1983
83. *Automot. Eng. (London),* Vol 7 (No. 3), June-July 1982, p 25-27
84. M.W. Toaz and M.D. Smalc, Clevite's Advanced Technology for the Diesel Piston, *Diesel Prog. North Am.,* Vol 51 (No. 6), June 1985, p 56
85. R.R. Bowles, D.L. Mancini, and M.W. Toaz, Design of Advanced Diesel Engine Pistons Employing Ceramic Fiber Reinforcement, *Seminar--Advanced Composites: The Latest Developments,* Nov 1986, American Society for Metals, p 21-27
86. T. Donomoto, K. Funatani, N. Miura, and N. Miyake, "Ceramic Fiber Reinforced Piston for High Performance Diesel Engines," TP 830252, Society of Automotive Engineers, 1983
87. T. Suganuma and A. Tanaka, Applications of Metal Matrix Composites to Diesel Engine Pistons, *J. Iron Steel Inst. Jpn.,* Vol 75 (No. 9), Sept 1989, p 1790-1797
88. R. Keribar, T. Morel, and M.W. Toaz, "An Investigation of Structural Effects of Fiber Matrix Reinforcement in Aluminum Diesel Pistons," TP 900536, Society of Automotive Engineers, 1990
89. T. Hayashi, H. Ushio, and M. Ebisawa, "The Properties of Hybrid Fiber Reinforced Metal and Its Application for Engine Block," TP 890557, Society of Automotive Engineers, 1989
90. R. Perrot, Engine Sleeves Made in a Hypereutectic Aluminum-Silicon Alloy by Powder Metallurgy, *European Symposium on Powder Metallurgy* (Stockholm), Vol 2, 1978, p 212-217
91. R. Perrot, Engine Cylinder Liners in High Silicon Aluminum Alloy Produced by Powder Metallurgy, *Rev. Alum.,* Mar 1979, p 129-135
92. N. Kuroishi, Y. Odani, and Y. Takeda, High Strength High Wear Resistance Aluminum-Silicon P/M Alloys, *Met. Powder Rep.,* Vol 40 (No. 11). Nov 1985, p 642-645
93. J.L Estrada and J. Duszczyk, Characteristics of Rapidly Solidified Al-Si-X Powders for High Performance Applications, *J. Mater, Sci.,* Vol 25 (No. 2A), Feb 1990, p 886-904
94. K. Hummert and V. Arnhold, Possibilities of Powder Metallurgy Aluminum Materials in Motor Construction, *3rd International Symposium on Aluminum + Automobile,* 1988, p 171-175
95. "High Strength Aluminum Alloy Powder Having Heat and Wear Resistance: Contains Silicon, Nickel, Iron and Manganese, and is Used for Cylinder Liners," U.S. patent 4,938,810
96. T. Hayashi, Y. Takeda, K. Akechi, and T. Fujiwara, "Rotary Car Air Conditioner Made with P/M Al-Si Wrought Alloys," TP900407, Society of Automotive Engineers, 1990
97. W. Kahl and J. Leupp, "High Performance Aluminum Produced by Spray Deposition," BNF 7th International Conference Material through the 90s
98. F. Hehermann *et al.,* "Comparison of Processing Atomized and Spray-Deposited Al-20 Si-5-Fe-2 Ni (wt.%) Aimed at Automotive Applications," PM90, Institute of Metals, London
99. J. Duszczyk *et al.,* The Osprey Preform Process and the Application to Aluminum Alloys Including Wear and High Temperature Resistance Al-20 Si-Cu Alloys, *Modern Developments in Powder Metallurgy,* Vol 18-21, Metal Powder Industries Federation, 1988, p 441-453
100. D.A. Parker, Material Innovation and Its Impact on the Development of Components for Reciprocating Engines, *Int. J. Veh. Des.,* IAVD Congress on Vehicle Design and Components, 1985, Inter-science Enterprises, Ltd., United Kingdom, p B35.1-30
101. A.E. Ostermann, "Experiences with Nickel-Silicon Carbide Coatings in Cylinder Bores of Small Aluminum Engines," TP 790843, Society of Automotive Engineers, 1979

102. J.D. Destefani, Ion Implantation Update, *Adv. Mater. Proc.,* Oct 1988, p 39-43
103. C.M. Preece and J.K. Hirvonen, *Ion Implantation Metallurgy,* TMS-AIME, 1980
104. I.L. Singer, Tribomechanical Properties of Ion Implanted Metals, *Ion Implantation and Ion Beam Processing of Materials,* MRS Symposia Proceedings, Materials Research Society, Vol 27, 1983, p 585-595
105. J.H. Hirvonen, Industrial Applications of Ion Implantation, *Ion Implantation and Ion Beam Processing of Materials,* MRS Symposia Proceedings, Materials Research Society, Vol 27, 1983, p 621-629
106. G.L. Kutner, Thermal Spray by Design, *Adv. Mater. Proc.,* Oct 1988, p 63
107. M.L. Thorpe, Thermal Spray Applications Grow, *Adv. Mater. Proc.,* Oct 1988, p 69
108. E.J. Kubel, Jr., Thermal Spraying Technology: From Art to Science, *Adv. Mater Proc.,* Dec 1987, p 69
109. C.W. Draper, Laser Surface Alloying: The State of the Art. *J. Met.,* June 1982, p 24
110. J.M. Rigsbee, Surface Engineering by Laser and Physical Vapor Deposition Techniques, *J. Met.* Aug 1984, p 31
111. E. Blank, M. Pierantoni, and M. Carrard, Laser Surface Modification of Al-Si Cast Alloys for Wear Resistance and Surface Toughness, *Surface Modification Technologies III,* 1990, T.S. Sudarshan and D.G. Bhat, Ed., TMS-AIME, 1990, p 493-510

Friction and Wear of Cemented Carbides

Henry J. Scussel, GTE Valenite

Introduction

CEMENTED CARBIDES, best known for their superior wear resistance, have a range of industrial uses more diverse than that of any other powder metallurgy product. Common uses include metalworking tools, mining tools, and wear-resistant components. All of these applications have one physical property requirement in common: the ability to resist wear. The variety of wear mechanisms encountered in service requires the use of a number of carbide grades with different chemical and metallurgical properties.

This article will discuss the raw materials used in the production of cemented carbides; the manufacturing methods employed; their physical, mechanical, and thermal properties; and the wear mechanisms encountered in service. Emphasis is placed on tungsten carbide-cobalt (WC-Co) or tungsten carbide-nickel (WC-Ni) materials as used in nonmachining applications.

Acknowledgements

The author gratefully acknowledges the assistance of R. James Franz in preparation of the photomicrographs, and the critiques by Steve J. Burden and Les J. Kastura of GTE Valenite, and J. Gary Baldoni and Steve T. Wayne of GTE Labs, Inc.

Raw Materials

The unique character of cemented carbides begins with the raw materials utilized in their manufacture. Table 1 lists the basic physical and mechanical properties of commonly used carbide and metallic binder materials. A more complete discussion of raw materials can be found in the article "Cemented Carbides" in *Properties and Selection: Nonferrous Alloys and Special-Purpose Materials,* Volume 2 of *ASM Handbook* (1990).

Table 1 Properties of refractory metal carbides and binder materials

Material	Hardness, HV (50 kg)	Crystal structure	Melting temperature		Theoretical density, g/cm^3	Modulus of elasticity		Thermal expansion,
			°C	**°F**		**GPa**	**10^6psi**	μm/m · K
Carbide								
TiC	3000	Cubic	3100	5600	4.94	451	65.4	7.7
VC	2900	Cubic	2700	4900	5.71	422	61.2	7.2
HfC	2600	Cubic	3900	7050	12.76	352	51.1	6.6
ZrC	2700	Cubic	3400	6150	6.56	348	50.5	6.7
NbC	2000	Cubic	3600	6500	7.80	338	49.0	6.7
Cr_3C_2	1400	Orthorhombic	1800	3250	6.66	373	54.1	10.3
WC	(0001) 2200 (1010) 1300	Hexagonal	2800	5050	15.63	696	101.0	(0001) 5.2 (1010) 7.3
Mo_2C	1500	Hexagonal	2500	4550	9.18	533	77.3	7.8
TaC	1800	Cubic	3800	6850	14.50	285	41.3	6.3
Binders								
Co	<100	Cubic/hexagonal	1495	2725	8.9	207	30.0	16.0
Ni	<100	Cubic	1455	2650	8.9	207	30.0	15.0

Tungsten carbide (WC) is manufactured through the reduction of tungsten oxide and subsequent carburization at 1400 to 1500 °C (2550 to 2730 °F). Particle sizes range from 0.5 to 30 μm. Each particle is composed of numerous tungsten carbide crystals. The tungsten powder is sometimes doped with small (<1 wt%) amounts of vanadium, chromium, or tantalum/niobium before carburization. These materials act as grain-growth inhibitors, particularly in the very fine (<1 μm) particle sizes.

Cobalt (Co) is the most widely used binder in WC-base hardmetals. Cobalt is the preferred binder due to its outstanding wetting and adhesion characteristics. As shown in Table 1, cobalt has a low-temperature hexagonal phase and as high-temperature cubic phase, with a phase transition at about 415 °C (780 °F). Cobalt is manufactured through reduction of cobalt oxides or derived from organic salts--particularly cobalt oxalate. The cobalt binder phase is altered significantly during milling with WC and subsequent liquid-phase sintering operations.

Nickel (Ni) is used as a binder in less than 10% of total carbide production because of poor WC wettability, which results in decreased hardness and toughness relative to cobalt grades at identical binder levels. Tungsten carbide-nickel grades offer slightly improved corrosion and oxidation resistance over cobalt binder grades.

Tantalum/titanium/niobium carbides (TaC/TiC/NbC) are used predominately in metal forming applications when metal pick up or galling of dies is problem. Tantalum carbide, in particular, is used in small quantities (<1 wt%) as a grain-growth inhibitor, and in fairly large amounts (>6 wt%) to provide increased hot hardness in metal cutting.

Chromium carbide is added to both WC-Co and WC-Ni grades is small quantities (<5 wt%) to improve corrosion and oxidation resistance. The WC-Ni grades alloy with chromium or chromium carbide have shown significantly improved corrosion resistance over conventional WC-Co or WC-Ni grades. Chromium or chromium carbide additions increase the tendency to from a brittle carbon-deficient ηphase, which can result in decreased strength and toughness.

Manufacturing Methods

Grade Powders are produced by combining WC with cobalt or nickel binder and, depending on the application, varying amounts of TaC/TiC/NbC. The grade powder is milled in conventional ball mills, attritor mills, or vibratory mills. The milling process reduces the particle size of the raw materials and also provides uniformity of mixture. Milling operations are typically carried out in a protective solvent such as alcohol, used to minimize heating and subsequent oxidation of the powder, and to disperse the powder particles to achieve intimate mixing. During the grade powder manufacturing process, 2 to 3 wt% of a solid lubricant such as paraffin wax is added. This lubricant reduces the potential for oxidation and provides green strength to as-pressed components. The lubricant/solvent/powder slurry is then dried to remove the solvent. Spray drying is the most widely used atomized through a nozzle and sprayed into a stream of nitrogen gas. The solvent is vaporized, condensed outside the chamber, and reused. The dried powder is now in the form of free-flowing spherical aggregates on the order of 150 to 250 μm in diameter (Fig. 1).

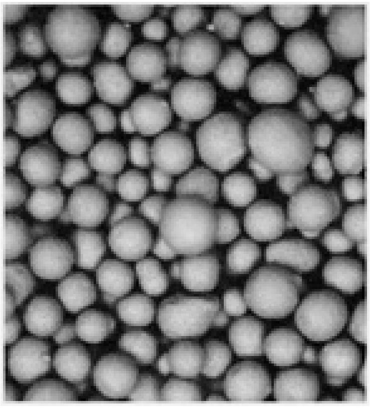


Fig. 1 Spray-dried cemented carbide powder. 40×

Pressing or powder consolidation processes include a number of vastly different techniques. Many wear components are produced on automatic or semiautomatic presses at pressures of 50 to 150 MPa (7 to 22 ksi). Press tooling typically consists of carbide dies, punches, and core rods to minimize wear. In cold isostatic pressing, the grade powder is consolidated into a rough billet or ingot using equal pressure from all directions. This rough billet is then preformed or machined to a net shape. Extrusion of carbide grade powders is used to produce long components of small, constant cross section. Higher lubricant levels ad different lubricant types are used in extrusion. There has been limited injection molding of cemented carbides.

Preforming or shaping operations are used to machine components to a net shape using abrasive (diamond) grinding wheels or single-point diamond tools. These operations are used when the final part cannot be pressed to its final shape, or the production quantity is too low to justify te investment in press tooling.

Sintering operations are carried out in batch-type or semicontinuous furnaces in either a vacuum hydrogen, or other inert atmosphere. A 400 to 500 °C (750 to 930 °F) hold dewaxes the parts, and the vaporized lubricant is condensed outside the heating chamber and discarded. Final sintering takes place at 1300 to 1600 °C (2370 to 2910 °F); the precise temperature depends on the cobalt content--the grades with the higher cobalt contents have the lower sintering temperatures. This final sintering temperature is above the eutectic temperature of the carbide-binder system, and the binder partially melts. The excellent wettability of WC by cobalt results in rapid liquid-phase sintering, which promotes coalescence of the WC particles and produces a fully dense, virtually porosity-free microstructure. Linear shrinkage on the order of 15 to 25% takes place.

The advantages of hot isostatic pressing (HIP) have been exploited since the early 1970s. Components requiring high reliability and/or surface integrity are HIPed to eliminate residual porosity, pits, or flaws. Materials are heated to a temperature above the liquidus, and the vessel is pressurized with an inert gas to sightly less than 100 MPa (15 ksi). The combination of pressure and temperature forces the binder into any residual pits or porosity. The result is porefree component with higher reliability. Recent advances have combined liquid-phase sintering and HIP into a single sinter-HIP process. Sinter-HIP uses lower pressures and higher temperatures than conventional HIP with no sacrifice in component reliability. The resulting sinter-HIP microstructure is more uniform than that produced by conventional HIP, and sinter-HIP is more cost effective (Ref 1).

Finishing operations include grinding with diamond wheels, electrical discharge machining (EDM) using wire or shaped electrode, and edge honing using a variety of abrasive techniques. Final lapping to mirrorlike finishes is accomplished using diamond-containing slurries or pastes.

Physical or chemical vapor deposition (PVD or CVD) are now used on the majority of metal cutting inserts and on some wear parts. The coatings are on the order of 5 μm thick and consist of titanium carbide (TiC), titanium nitride (TiN), aluminum oxide (Al_2O_3), or a combination thereof. Their purpose is to minimize the wear process during steel machining; in particular, to minimize the dissolution of the workpiece material into the cutting tool. Titanium nitride coatings are also purported to reduce the frictional forces at the tool/workpiece interface.

The methods of application for CVD and PVD coatings differ substantially. The CVD coatings are applied at temperatures of about 1000 °C (1830 °F), and as a result contain higher levels of residual stresses due to the difference in both thermal expansion coefficients and elastic moduli between the substrate and the coating. This results in a significant decrease (30%) in three-point bending strength of coated parts compared to uncoated parts. The PVD process is applied at less than 500 °C (930 °F) and results in reduced stresses in the coating and minimizes any loss in strength. Coatings applied by PVD or CVD do not have the same utility in nonmachining applications where high-temperature wear resistance is not required. Coatings may provide some improvement in corrosion resistance or resistance against smearing or pickup of the workpiece onto a forming tool.

Properties

The properties of cemented carbide grades are predominately determined by their chemical composition and the grain size of the tungsten carbide in the sintered part. Table 2 summarizes the properties of some typical carbide grades used in wear part application. Table 3 relates these compositions to application areas.

Table 2 Properties of representative cobalt-bonded cemented carbide grades

Nominal composition	Grain size	Hardness, HRA	Density, g/cm^3	Transverse strength		Compressive strength		Modulus of elasticity		Relative abrasion resistance[(a)]	Coefficient of thermal expansion, μm/m · K		Thermal conductivity, W/m · K
				MPa	ksi	MPa	ksi	GPa	10^6 psi		at 200 °C (390 °F)	at 1000 °C (1830 °F)	
97WC-3Co	Medium	92.5-93.2	15.3	1590	230	5860	850	641	93	100	4.0	. . .	121
94WC-6Co	Fine	92.5-93.1	15.0	1790	260	5930	860	614	89	100	4.3	5.9	. . .
	Medium	91.7-92.2	15.0	2000	290	5450	790	648	94	58	4.3	5.4	100
	Coarse	90.5-91.5	15.0	2210	320	5170	750	641	93	25	4.3	5.6	121
90WC-10Co	Fine	90.7-91.3	14.6	3100	450	5170	750	620	90	22	. . .	. . .	. . .
	Coarse	87.4-88.2	14.5	2760	400	4000	580	552	80	7	5.2	. . .	112
84WC-16Co	Fine	89	13.9	3380	490	4070	590	524	76	5	. . .	. . .	. . .
	Coarse	86.0-87.5	13.9	2900	420	3860	560	524	76	5	5.8	7.0	88
75WC-	Medium	83-85	13.0	2550	370	3100	450	483	70	3	6.3	. . .	71

(a) Based on a value of 100 for the most abrasion-resistant material

Table 3 Nominal composition and properties of representative cemented carbide grades and their applications

Typical application	Binder content, wt%	Grain size	Hardness, HRA
Heavy blanking punches and dies, cold heading dies	20-30	Medium	85
Heading dies (severe impact), hot forming dies, swaging dies	11-25	Medium to coarse	84
Back extrusion punches, hot forming punches	11-15	Medium	88
Back extrusion punches, blanking punches and dies for high shear strength steel	10-12	Fine to medium	89
Powder compacting dies, Sendzimir rolls, strip flattening rolls, wire flattening rolls	6	Fine	92
Extrusion dies (low impact), light blanking dies	10-12	Fine to medium	90
Extrusion dies (medium impact), blanking dies, slitters	12-16	Medium	88
Corrosion-resistant grades, valves and nozzles, rotary seals, bearings	6-12	Fine to medium	92
Corrosion-resistant grade with good impact resistance for valves and nozzles, rotary seals and bearings	6-10 Ni	Medium	90
Deep draw dies (nongalling), tube sizing mandrels	10 Co with TiC and TaC	Medium	91

Hardness is typically measured on the Rockwell A scale with values ranging from 83.0 HRA for high-cobalt coarse-grain grades, to 93.0 HRA for low-cobalt fine-grain grades. Vickers diamond pyramid hardness (HV) is widely used in Europe; values range from 800 to 2000 kg/mm^2 using a 30 kg load. The precision and accuracy of hardness testing is influenced significantly by the surface finish of the testpiece, parallelism between top and bottom surfaces, and the quality of hardness standards and diamond penetrators. For straight WC-Co grades (those not containing TiC, TaC, or similar additions) with comparable WC grain size, hardness decreases with increasing binder content. Figure 2(a) illustrates the relationship between hardness and cobalt content/WC grain size.

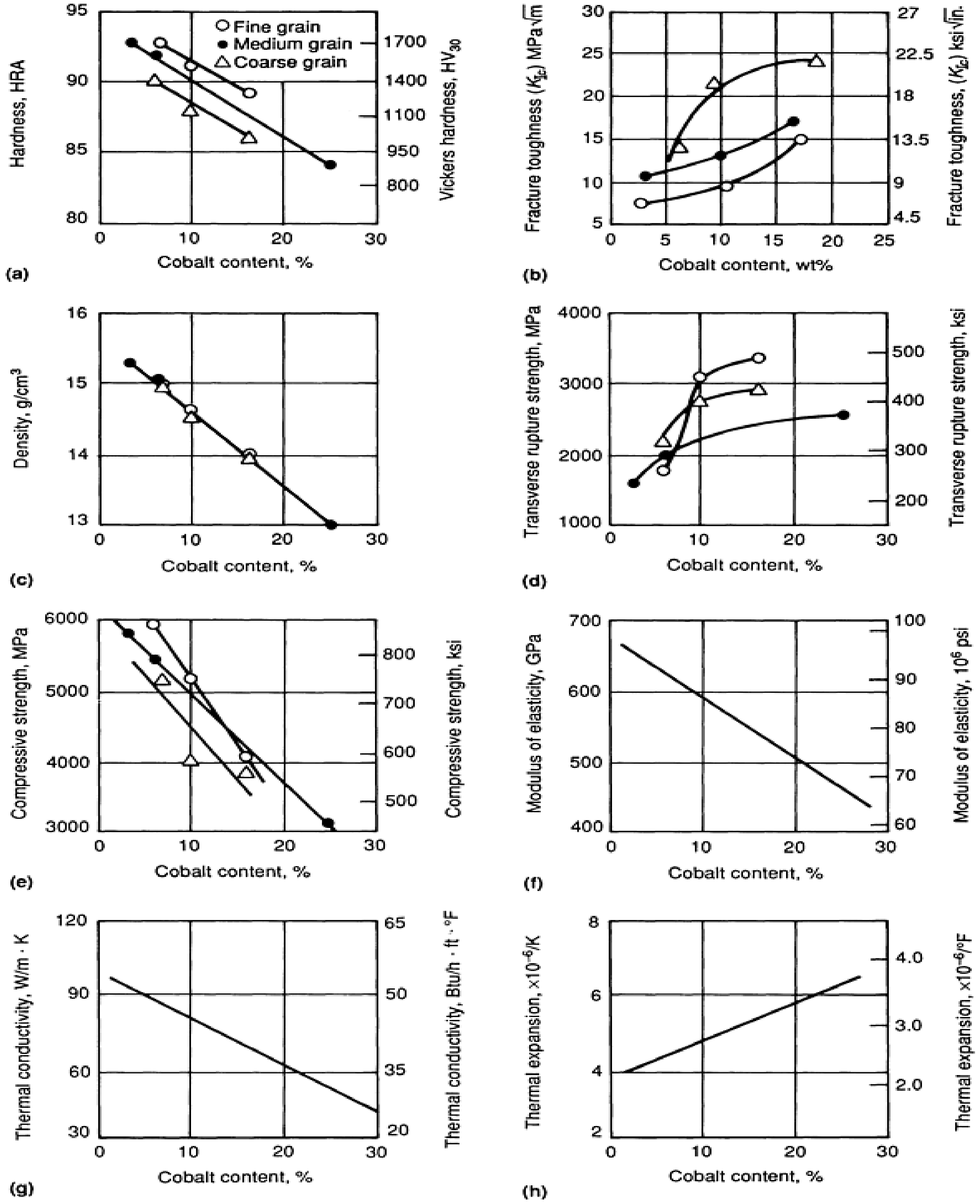


Fig. 2 Variation of properties with cobalt content and grain size for unalloyed grades of cemented carbide

Fracture toughness (K_{Ic}) values indicate the resistance of a material to fracture from intrinsic flaws. A variety of test methods and specimen geometries are used, so caution must be exercised when comparing reported values from different manufacturers. Fracture toughness increases with both increased cobalt content or WC grain size, as shown in Fig. 2(b).

Density varies inversely with cobalt content, as shown in Fig. 2(c). Porosity levels also influence density. Cemented carbide grades used in ferrous alloy machining applications contain higher amounts of TiC/TaC and have density values from 10 to 14 g/cm^3.

Transverse rupture strength (TRS), or three-point bending strength, is the most common method of determining the fracture strength of cemented carbides. Rectangular samples(5 × 6 × 19 mm, or 0.2 × 0.25 × 0.75 in.) are loaded as shown in Fig. 3. The TRS is then calculated using:

$$\text{TRS} = \frac{3FL}{WH^2}$$

where F is the load at fracture, L is the span between supports, and W and H are the width and height of the test bar, respectively. The test itself is sensitive to test bar size (several variations are in use), surface finish, and other test parameters. The response of TRS to cobalt content and tungsten carbide grain size is shown in Fig. 2(d).

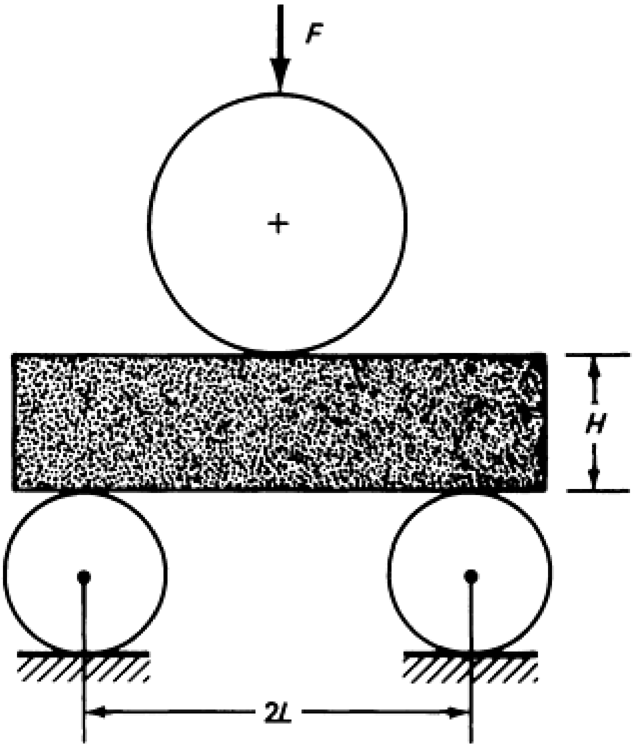


Fig. 3 Schematic of transverse rupture strength testing (three-point bending)

Compressive strength of cemented carbides is greater than that of almost any group of materials, metallic or nonmetallic. Uniaxial compressive strength tests are performed using straight cylindrical samples, a cylinder with reduced diameter at the center of the part to localize fracture, or a straight cylinder held within a sleeve. Reported compressive strength values can vary significantly depending on the size and geometry of the test specimen. Compressive strength varies inversely with cobalt content, and for a given cobalt content, finer grain sizes give the highest value. Figure 2(e) summarizes this compressive strength versus cobalt content/grain size relationship.

Modulus of elasticity, or Young's modulus, also varies inversely with cobalt content (Fig. 2f), but is independent of WC grain size (Ref 2). The elastic modulus of WC is higher than that of any other commercially available material except diamond and cubic boron nitride. As a result, WC-Co alloys have elastic moduli 2 to 3 times those of cast irons or steels.

Elastic modulus measurements are carried out using either mechanical or sonic methods. Mechanical methods involve loading a WC-Co beam and measuring the amount of deflection. Sonic methods are more widely used and utilize resonant vibration of a cylindrical or square rod. The resonant frequency depends on the dimensions of the testpiece, the density of the material, and the elastic modulus of the material.

Thermal conductivity of WC-Co alloys is important in machining applications because the ability of the tool to conduct heat away from the tool/workpiece interface has a definite effect on tool performance. In nonmachining applications, such as a rotary mechanical-pump seal, the tungsten carbide seal ring must have high thermal conductivity to ensure heat flow away from the rotary seal/stationary seal interface. Thermal conductivity decreases with increasing cobalt content and is unaffected by WC grain size (Ref 3), as illustrated in Fig. 2(g). The additional of TiC reduces the thermal conductivity significantly.

Coefficients of thermal expansion are an important design consideration when using WC-Co materials. The linear coefficient of thermal expansion of WC-Co increases with increasing cobalt content (Fig. 2h), and is independent of grain size. Typical low-carbon steels, tool steels, and stainless have thermal expansion coefficients 2 to 3 times greater than those of carbides. In metal forming applications at elevated temperatures, such as warm forming or extrusion, this difference, must be taken into consideration when designing steel/carbide assemblies. This expansion coefficient mismatch also complicates brazing operations when joining cemented carbides to metals.

Porosity in carbides is typically less than 0.1 vol%. ASTM procedure B 276 classifies porosity into three types:

- A-type: pores less than 10 μm in diameter
- B-type: pores between 10 and 25 μm in diameter
- C-type: porosity caused by the presence of uncombined carbon

Porosity is evaluated by observing a polished, unetched surface, and comparing it with standards supplied in the ASTM procedure.

Microstructures for nine typical WC-Co grades are shown in Fig. 4. The microstructures are characterized by the angular, sharp WC grains surrounded by the cobalt binder. Tungsten carbide grain sizes as shown in Fig. 4 vary from <1 μm in the fine grain sizes to 8 to 10 μm in the coarse grain size grades. Microstructures of grades for use as cutting tools and anomalies such as graphite and the carbon-deficient η phase are illustrated in the article "Cemented Carbides" in *Properties and Selection: Nonferrous Alloys and Special-Purpose Materials,* Volume 2 of *ASM Handbook* (1990).

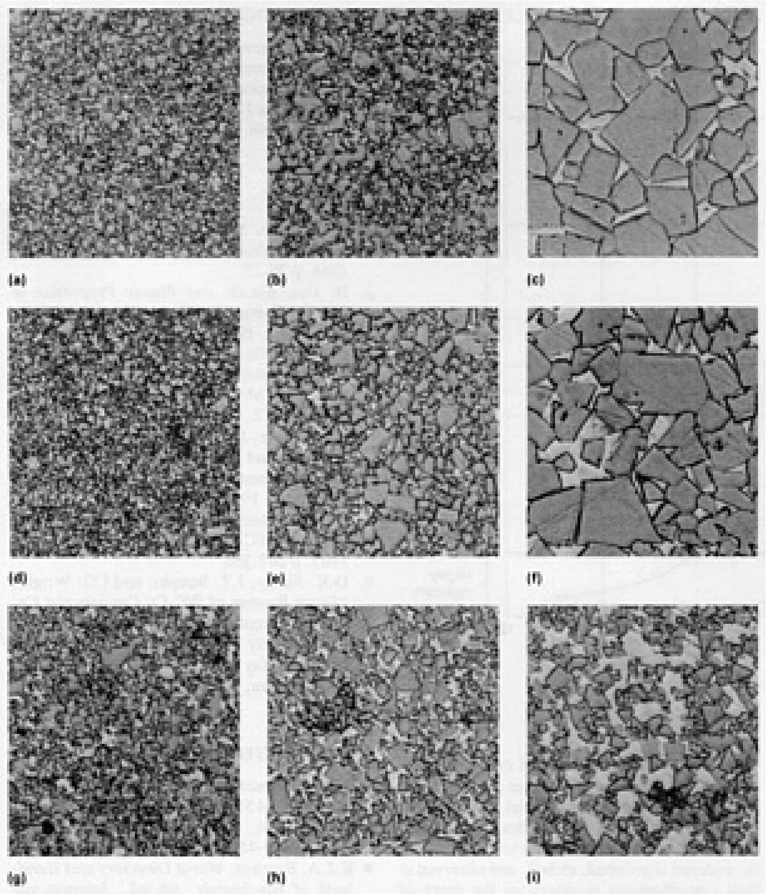


Fig. 4 Cemented carbide microstructures. (a) 94WC-6Co, fine grain size. (b) 94WC-6Co, medium grain size. (c) 93WC-7Co, coarse grain size. (d) 90WC-10Co, fine grain size. (e) 90WC-10Co, medium grain size. (f) 90WC-10Co, coarse grain size. (g) 84WC-16Co, fine grain size. (h) 84WC-16Co, medium grain size. (i) 75WC-25Co, medium grain size. All at 1500×, 3 min etch

Wear Properties of Cemented Carbides

Superior abrasive wear resistance is probably the major reason for the selection of cemented carbides in a wide variety of industrial applications. This superior wear resistance can generally be attributed to their unique composition, which consists of 80 to 95% hard, wear-resistant, fine WC grains combined with a cobalt binder that provides a small amount of ductility to the material. Studies of carbides have shown that abrasive wear involves rounding, fragmentation, and pullout of the WC grains, and subsequent removal of the soft binder phase (Ref 4, 5).

ASTM test B 611 specifies a method for determining the abrasive wear resistance of carbides. In this test, a rectangular carbide testpiece is held against a rotating wheel (either steel or rubber) for a fixed number of revolutions. The test is run either wet or dry. A schematic of the test apparatus is shown in Fig. 5. A steady stream of 30 mesh alumina sand is introduced directly into the carbide/wheel interface. The volume of the testpiece (measured in cubic centimeters) is determined both before the test, and after the predetermined number of revolutions. The volume loss is inversely related to the wear resistance. Typical volume losses for low-cobalt, fine-grain materials is on the order of 3 mm^3. A D2 tool steel is often used as reference material in this test and exhibits a volume loss of 40 to 45 mm^3. Materials are ranked using either the reciprocal of the volume loss (low volume loss equates to high wear resistance), or using a particular grade (usually a WC-6Co grade with hardness of 92.0 HRA) as a reference, and reporting the performance of the grade as a percentage of this standard. Some caution must be exercised when comparing values reported by different manufacturers. Cemented carbide manufacturers have not agreed to a single test method, and there will be variation in the values and the units used to report wear resistance.

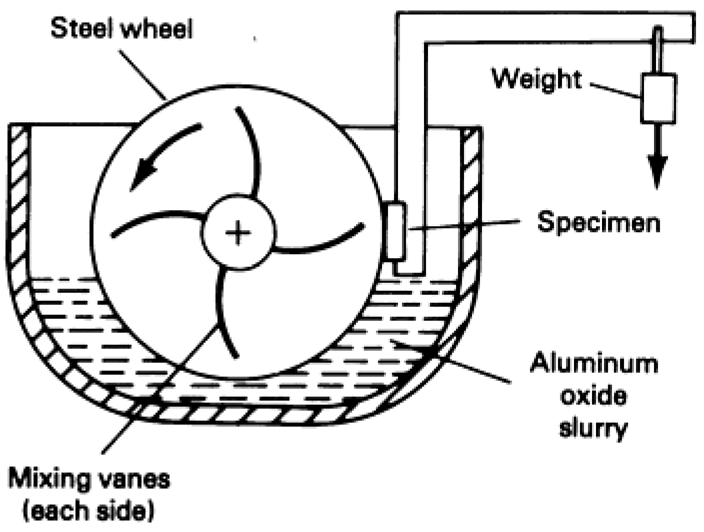


Fig. 5 Schematic of abrasive wear resistance apparatus

Another method involves a pin-on-disk apparatus (Ref 4). In this test, a resin-bonded diamond-covered disk is rotated at approximately 40 rev/min. A load is applied through a square WC-Co sample that is swept and rotated against the diamond wheel. The volume of the sample is determined before and after, and the volume loss is determined. Values are reported in a manner similar to that of the ASTM procedure above.

Figure 6 shows the relationship among fracture toughness, hardness, WC grain size, and abrasive wear resistance (Ref 4). The grades shown as 6C, 6M, 6F, 12C, 12M, and 12F represent 6 and 12% cobalt grades with coarse, medium, and fine grain size, respectively. The abrasive wear resistance increase with decreasing cobalt content and decreasing grain size. The abrasive wear resistance appears to be influenced more strongly by the WC grain size than by the cobalt content.

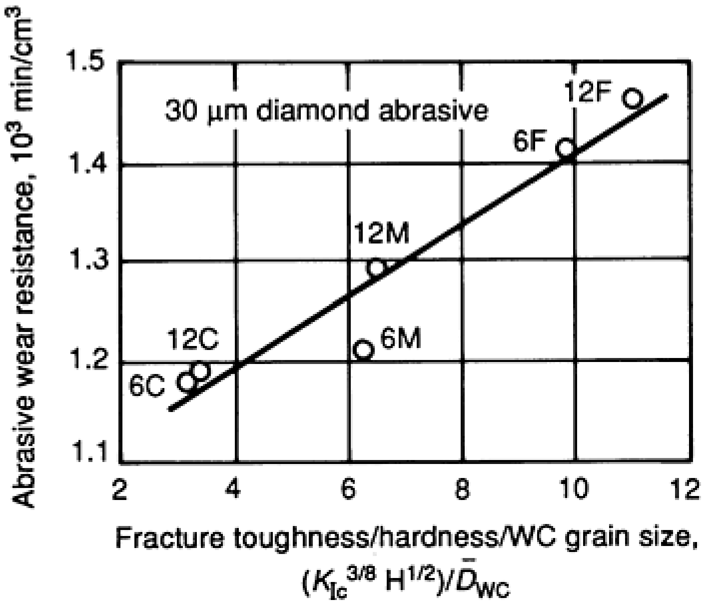


Fig. 6 Results of abrasive wear resistance tests. See text for description of data points. Source: Ref 4

The results of an abrasion test cannot be used to described the performance of a carbide grade when anything other than pure abrasive wear is present. For example, when cratering or dissolution wear is taking place during machining

operations, abrasion resistance test results are not meaningful. The same can be applied to metal pickup or galling during metal forming. In summary, abrasive wear resistance test results can only be applied to those instances when abrasion is the dominant wear mechanism.

Erosion resistance of carbides is important in applications such as sand blast-spray nozzles, seals in slurry pumps, and component parts in the oil industry. The success of carbides as an erosion-resistant material is again due to its unique composite structure of wear-resistant WC particles in a ductile cobalt matrix (Ref 4, 6).

The erosion resistance test involves impacting the WC-Co surface with 50 μm alumina particles at over 100 m/s (325 ft/s) for a set period of time under an inert gas carrier. The volume of the resultant crater is determined and reported as a volume loss. The smaller crater volumes equate to improved erosion resistance. Figure 7 illustrates a test apparatus for gas stream alumina particle erosion tests.

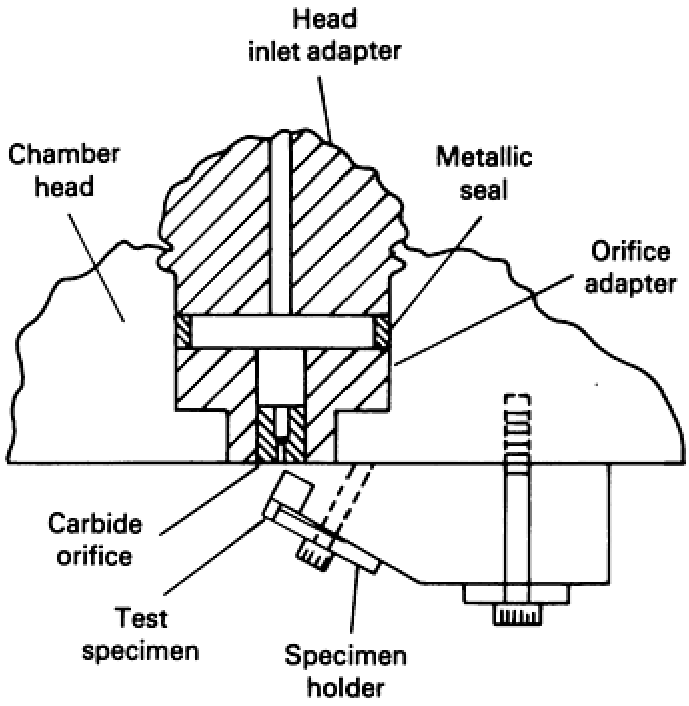


Fig. 7 Schematic of specimen fixturing for particle erosion testing. Source: Ref 6

The performance of carbides in erosion tests closely parallels their abrasive wear performance: the erosion resistance increases with decreasing cobalt content. The effect of WC grain size is not as strong as in straight abrasion, but the erosion rate does decrease slightly with increasing carbide grain size for cobalt percentages greater than 15%. Figure 8 shows the relationship between crater volume and the cobalt volume fraction. This relationship suggests that a binderless carbide (<5 vol%) would be preferred for erosion applications. However, grades with less than 5 vol% binder tend to exhibit higher wear rates due to the presence of increased grain boundary porosity after consolidation (Ref 6). These binderless alloys may not have the fracture toughness required to handle the thermal or mechanical loading that will be encountered in service. As such, design engineers may be required to compromise some erosion resistance for improved structural integrity and mechanical reliability.

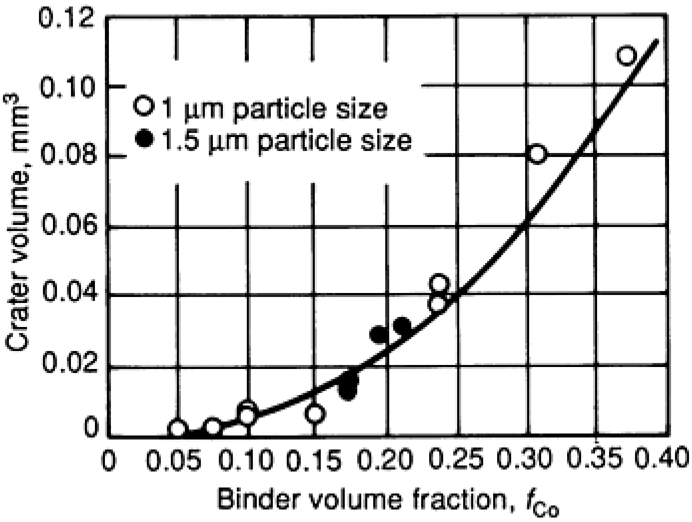


Fig. 8 Results of particle erosion tests. Source: Ref 6

Corrosion resistance is not typically thought of as a requirement for carbides, and in terms of straight corrosion resistance, there are commercially available materials that are clearly superior. However, when a combination of corrosion resistance, wear resistance, stiffness, toughness, and thermal conductivity is required--such as for rotary mechanical seals in industrial pumps--cemented carbides are the material of choice.

The corrosion resistance of carbides is limited by the susceptibility of the cobalt binder to chemical attack, although there are some corrosive media that attack WC. The corrosive media typically dissolve the cobalt binder from the matrix, leaving behind a weak, unsupported skeleton of tungsten carbide grains, which are easily abraded away. The corrosion resistance of straight WC-Co alloys is, in general, inversely related to that of the binder content. The straight substitution of nickel binder for cobalt does provide limited improvement in both corrosion and oxidation resistance. There are several grades available that utilize chromium additions to the cobalt binder to improve corrosion resistance. Chromium, however, promotes the formation of the carbon-deficient η phase with a resultant decrease in toughness and strength. The development of nickel binder grades alloyed with chromium, molybdenum, and other elements has resulted in a significant improvement in corrosion resistance with little sacrifice in strength or toughness. Figure 9 shows corrosion resistance versus pH for three different types of materials: WC-Co, alloyed WC-Ni, and TiC-Ni cermets (Ref 7). Although TiC-base cermets exhibit superior corrosion resistance, their use has been severely limited due to their inferior strength and lower thermal conductivity.

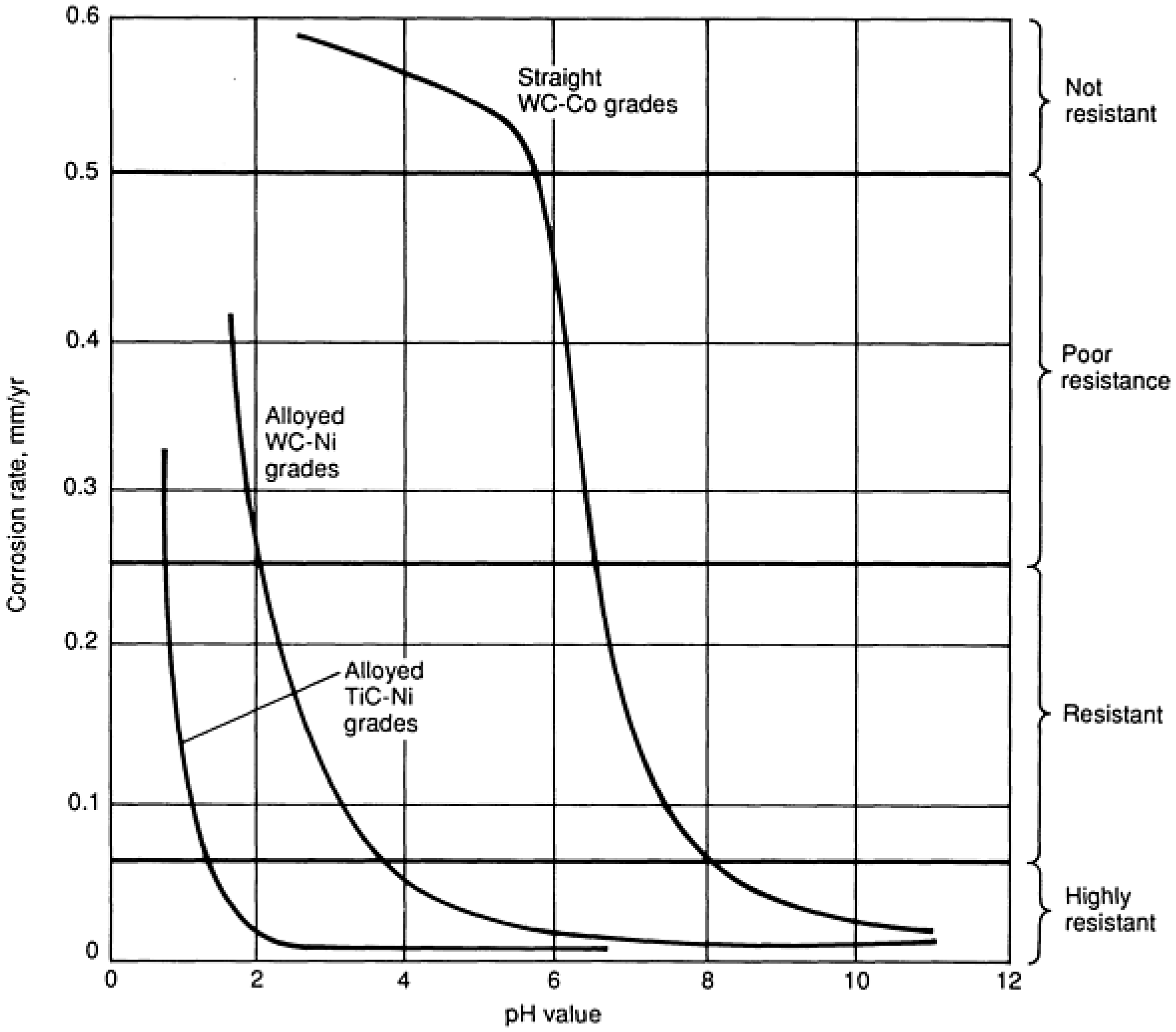


Fig. 9 Corrosion resistance of cemented carbides. Source: Ref 7

Corrosion resistance is determined using immersion methods. Test specimens are immersed in the media for a set period of time, at a constant temperature, and at a known pH level. The corrosion behavior of the material is determined by one of three methods. The first is a metallographic examination in which a cross section of the material is polished, etched, and observed at high magnifications to determine the depth of binder removal. The second involves tumbling the test specimen after immersion. Tumbling abrades the unsupported network of WC grains and leads to a final weight loss measurement. The third method utilizes atomic absorption techniques to measure the amount of binder (cobalt or nickel) that has been dissolved in the corrosive media. A low concentration of binder in the media indicates a high corrosion resistance. Additional information on the corrosive behavior of carbides can be found in the article "Corrosion of Cemented Carbides" in *Corrosion,* Volume 13 of *ASM Handbook* (1987).

References

1. R.C. Leuth, Moldless Hot Pressing of Cemented Carbides, *Carbide Tool J.,* Nov-Dec 1984, p 20-25
2. H. Doi, *Elastic and Plastic Properties of WC-Co Composite Alloys,* Freund Publishing (Israel), 1974, p 32-40
3. A. Perecherla and W. Williams, Room Temperature Thermal Conductivity of Cemented Transition Metal Carbides, *J. Am. Ceram. Soc.,* Vol 12, Dec 1988, p 1130-1133
4. S.F. Wayne, J.G. Baldoni, and S.T. Buljan, Abrasion and Erosion of WC-Co with controlled Microstructures, *Tribol. Trans.,* Vol 33 (No. 4), 1990, p 611-617

5. J. Larson-Basse, Binder Extrusion in Sliding Wear of WC-Co Alloys, *Wear,* Vol 105, 1985, p 247-256
6. D.K. Shetty, J.T. Stropki, and I.G. Wright, "Slurry Erosion of WC-Co Cermets and Ceramics," Preprint No. 84-AM-3A-3, American Society of Lubrication Engineers (ASLE), May 1984
7. L. Lindholm, U.S. patent 4,497,660, 5 Feb 1985

Selected References

- ASM Committee on Tooling Materials, Superhard Tool Materials, *Tool and Die Failure Source Book,* American Society for Metals, 1982, p 400-410
- K.J.A. Brookes, *World Directory and Handbook of Hardmetals,* 4th ed., International Carbide Data, 1987
- H.E. Exner, Physical and Chemical Nature of Cemented Carbides, *Int. Met. Rev.,* Vol 243 (No. 4), 1979, p 149-173
- A.T. Santhanam, P. Tierney, and J.L. Hunt, Cemented Carbides, Vol 2, *ASM Handbook,* ASM International, 1990, p 950-977
- V.K Sarin, Cemented Carbide Cutting Tools, *Advances in Powder Technology,* American Society for Metals, 1985, p 253-288
- P. Schwarzkopf and R. Kieffer, *Cemented Carbides,* Macmillan, 1960

Friction and Wear of Metal-Matrix Composites

P.K. Rohatgi, Y. Liu*, and S. Ray, University of Wisconsin-Milwaukee**

Introduction

FOR MOST OF THE PAST CENTURY, research and development in tribology have been directed at meeting the severe conditions of mechanical systems, such as advanced engines, that require increasingly high working temperatures and running speeds. The development of new materials that can meet these severe requirements has become a major imperative. Metal-matrix composites are attracting considerable interest worldwide because of their superior mechanical and tribological properties. These advanced materials have a metal matrix in which nonmetallic fibers, particles, or whiskers are dispersed. For use in tribological applications, metal-matrix composites must be able to support a load without undue distortion, deformation, or fracture during performance and to maintain controlled friction and wear over long periods without seizure under working conditions (Ref 1).

This article will review the structure, processing, and properties of those metal-matrix composites (both particulate and fiber reinforced) that have been studied for their tribological properties. In addition, it will discuss the current and potential uses of metal-matrix composites in applications where tribological properties are of paramount importance, such as electrical contact brushes, bearings, pistons, cylinder liners, and brake drums.

A variety of nonmetallic particles have been dispersed in different metal systems to develop metal-matrix nonmetallic particle composites; a selected list of these particles is given in Table 1 (Ref 2, 3, 4). These particles can be roughly divided into two different groups on the basis of hardness: (1) hard particles with hardness of 4 to 30 GPa (580 to 4350 ksi), such as SiC, Al_2O_3, and silica, and (2) soft particles with hardness below 2 GPa (290 ksi), such as graphite and MoS_2, which are primarily added for solid lubrication purposes. Particles with values between these ranges are considered to have intermediate hardness. This arbitrary division is based on the fact that composites containing hard particles generally exhibit different friction and wear behavior (because of different wear mechanisms) than composites containing soft particles in alloys with the same matrix.

Table 1 Density and hardness of selected particles for metal-matrix particle composites

Particle type	Density		Mohs hardness	Hardness	
	g/cm^3	**lb/in.^{3}**		**GPa**	**ksi**
TiC	4.93	0.178	9	20-30	2900-4350
SiC	3.22	0.116	9-9.5	24.5-29	3550-4205
Al_2O_3	3.97	0.143	9	18-26	2610-3770
WC	15.63	0.565	8-9	24	3480
TiN	5.22	0.189	. . .	20.61	2990
Si_3N_4	3.44	0.125	. . .	16-20	2320-2900
VC	5.77	0.208	. . .	16-18	2320-2610
TiO_2	4.26	0.154	6-6.5	12.7	1840
$ZrSiO_4$	4.56	0.165	7.5	11	1595
ZrO_2	5.89	0.213	. . .	6.5-10	940-1450
SiO_2	2.65	0.096	7	8	1160
MgO	3.55-3.68	0.128-0.133	5.5	. . .	. . .
Glass	2.48-2.54	0.090-0.092	4.5-6.5	4	580
Mica	2.7-2.8	0.098-0.10	2.8	0.3	43.5
Talc	2.58-2.83	0.093-0.102	1-2	. . .	. . .
MoS_2	4.62-4.8	0.167-0.173	1-1.5	. . .	. . .
Graphite	2.09-2.23	0.076-0.081	0.5-1	0.25	36

(a) Polytetrafluoroethylene

Particulate-reinforced composites cost less than fiber-reinforced composites, owing to the lower costs of the particles. In addition, the mechanical and physical properties of particle composites are generally isotropic. Cast metal-matrix particulate composites represent the lowest cost composites, and these have found the most tribological applications.

Table 2 lists a number of metal-matrix composites synthesized using the two types of particles, along with particle size, volume fraction, matrix alloys, and methods of fabrication. The matrix alloys used for the synthesis of composites include those based on aluminum, copper, silver, iron, magnesium, steel, titanium, and cobalt. Synthesis techniques include powder metallurgy, mechanical alloying, spray dispersion, squeeze casting, stir casting, compocasting, and low-pressure infiltration. In recent years, the use of inexpensive liquid metallurgy techniques, including stir casting, squeeze casting, and pressure infiltration, has become widespread in the production of metal-matrix particulate composites (Ref 5, 6, 7).

Table 2 Selected metal-matrix nonmetallic particle composites and methods used for their synthesis

Particle type	Dispersoid size, μm	Vol%	Matrix alloy	Method of fabrication
SiC	1-20	0.3-20	Al-Si, Al-Cu, Al-Cu-Mg	Vacuum slurry casting, squeeze casting, powder metallurgy
TiC	<40-212	8-40	Al-Cu, Al-Mg, Ti-Al-V, steel	Slurry casting, squeeze casting, powder metallurgy, laser melt-particle injection, casting
Al_2O_3 (bauxite), 87.9% Al_2O_3	0.01-200 <50	0.5-10 1-20	Al-Mg, Al-Cu, Al-Si, Cu, steel, Mg	Slurry casting, squeeze casting, powder metallurgy
WC	106-150	. . .	Ti-Al-V, Co-base	Laser melt-particle injection, powder sintering
M_7C_3 (Cr-rich)	18-38	. . .	Co-Cr	Powder metallurgy
ZrO_2/$ZrSiO_4$	5-80	1-4	Cu, Al, steel	Slurry casting, bottom pouring, spray dispersion, powder metallurgy
TiO_2/MgO	. . .	10	Cu, Al, steel	Slurry casting, bottom pouring, spray dispersion, powder metallurgy
Glass/SiO_2	30-110	2-10	Al-Mg, Cu	Slurry casting, bottom pouring, powder metallurgy
Mica/talc	40-180	3-10	Al-Cu-Mg, Ag, Cu-Sn	Slurry casting, compocasting, powder metallurgy
Shell char	125	15	Al-Si-Mg	Slurry casting, squeeze casting
Graphite	15-800	1-750	Al, Cu, Ag, iron	Slurry casting, squeeze casting, powder metallurgy
PTFE	. . .	20-40	Cu, Ag, Cu-steel	Powder metallurgy
MoS_2	0.5/5	1-80	Cu, Cu-Ta	Powder metallurgy
$MoSe_2$	. . .	20-80	Fe-Pb, Ag-Cu, Ag	Powder metallurgy

Notes

* Presently with National University of Singapore

** On leave from Department of Metallurgical Engineering, University of Roorkee, India

Mechanical Properties of Metal-Matrix Composites

The mechanical properties of a material are important in determining its tribological behavior. Wear can often be attributed to cracks that originate in the heavily deformed subsurface region and propagate to the surface, thereby releasing debris. Therefore, the deformation and fracture behavior of a material under both uniaxial and triaxial load assumes importance in the context of its tribological response.

In metal-matrix composites, mechanical properties depend on the amount, size, shape, and distribution of the dispersed phase apart from the mechanical properties of the matrix material and on the nature of the interface. By definition, a composite material generally requires an amount of dispersed phase (> 1 vol%) of a size (> 1 μm) that allows this constituent to be load bearing and not act merely to control the movement of dislocations, as in dispersion-strengthened materials. The shape of the dispersed phase is so important in determining its load-bearing capacity that composites have been classified on this basis: (1) fiber-reinforced composites with both continuous and discontinuous fibers and (2) particle- or whisker-reinforced composites. The aspect ratio (that is, the ratio of length to diameter of a fiber) generally characterizes the shape. In continuous-fiber composites, the load is applied directly to both the matrix and the fiber. In discontinuous-fiber composites or particle-reinforced composites, the load is transmitted to the dispersoid through the matrix. Mismatch of strain in the matrix and the dispersoid across the interface results in a shear stress at the interface. The transfer of load to the dispersed phase depends on the magnitude of the shear stress as described in simplistic "shear lag" models for discontinuous fibers and particles in a matrix (Ref 8, 9). If the shear stress that develops at the interface exceeds the strength of the interface, debonding will occur at the interface, and this may develop into a crack that can propagate. For a given condition of loading, the load shared by the dispersed phase increases with its aspect ratio.

Going beyond the framework of elasticity theory, the presence of hard dispersed particles will cause additional strain hardening; this may not be significant in the case of soft particles with a shear modulus lower than that of the matrix alloy. Thus, composites with soft particles such as graphite have lower strength compared with the matrix alloy, as shown in Fig. 1 (Ref 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25). The strength of composites that contain hard particles increases with the volume percentage of particles in the composite. However, the ductility of composites that contain both soft and hard particles decreases compared with that of the matrix alloy (Fig. 2) presumably because of debonding of the interface at low strain, as often indicated by serrations in tensile stress-strain diagrams (Ref 26).

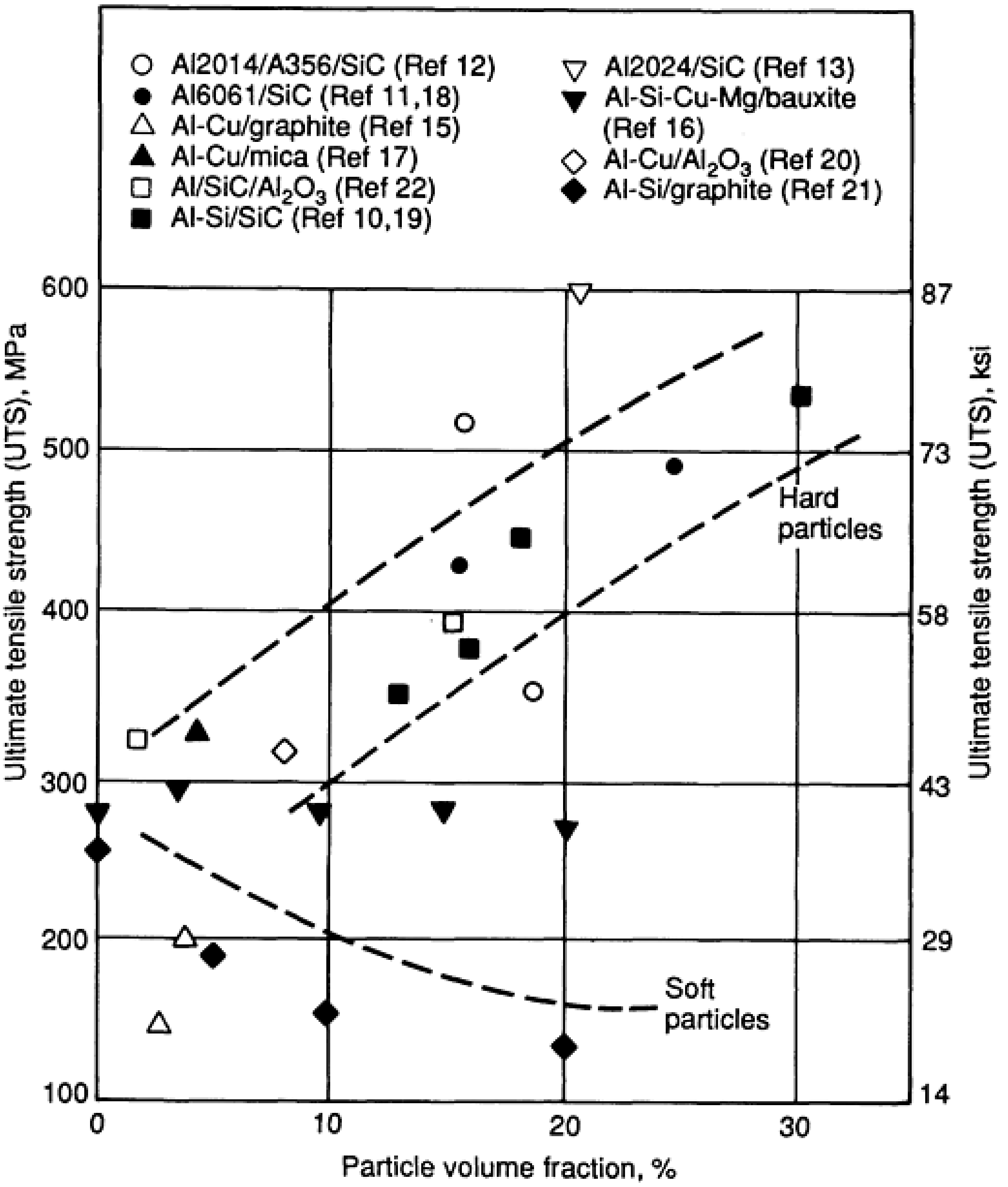


Fig. 1 Ultimate tensile strength as a function of particle volume fraction for aluminum alloy composites

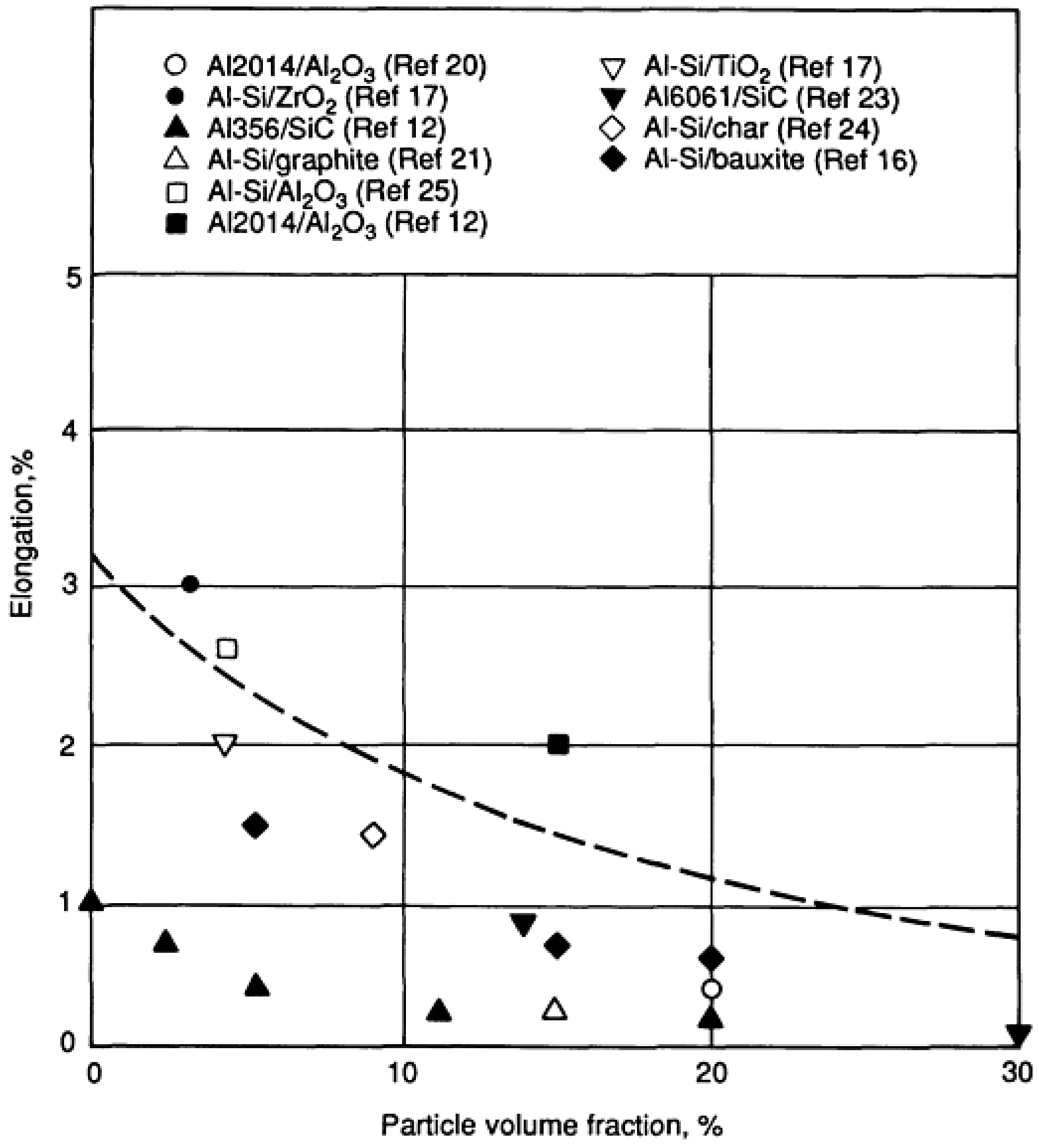


Fig. 2 Elongation as a function of particle volume fraction for aluminum alloy composites

The mechanical properties of discontinuous-fiber composites are similar to those of particle- or whisker-reinforced composites, except that the strength of discontinuous-fiber composites increases with aspect ratio. In the case of continuous-fiber composites, strength increases with volume fraction of the fibers (Fig. 3), even for soft fibers such as graphite (Ref 27), because these fibers are strong. The physical properties of the fibers used to reinforce metal-matrix composites are given in Table 3. The mechanical properties of fiber-reinforced metal-matrix composites are given in Table 4.

Table 3 Properties of fibers used in metal-matrix composites tested for tribological applications

Fiber	Composition	Diameter, μm	Density		Strength		Young's modulus	
			g/cm^3	**lb/in.^{3}**	**GPa**	**ksi**	**GPa**	**ksi**
Thornel 50	Carbon	6	1.67	0.060	2.40	348	413	59,900
Thornel 300	Carbon	7	1.75	0.063	2.65	385	227	32,920
Hm 3000	Carbon	7	1.90	0.069	2.43	352	365	52,940
Modmor 1	Carbon	8	1.85	0.067	2.05	297	317	45,970
Celion 6000	Carbon	7	1.75	0.063	2.92	423	228	33,070
VSA-11	Carbon	11	1.99	0.072	1.20	174	380	55,100

Saffil RF/Saflmax	δ-Al_2O_3	3	3.30	0.120	2.00	290	300	43,500
SN-N-X (whisker)	Si_3N_4	0.2	3.20	0.116	3.00	435	350	50,760
Tokamax (whisker)	SiC	0.5	3.20	0.116	10.00	1450	480	69,600
SiC whisker	β-SiC	0.1-1	. . .	. . .	3-10	435-1450	400-700	58,000-101,500
Fortafil 4R	Carbon	8	. . .	. . .	2.4 ± 0.6	348 ± 87	218 ± 30	31,600± 4,350
Hercules A	Carbon	8	. . .	. . .	3.5 ± 0.7	508 ± 100	243 ± 29	35,240 ± 4,200
Fiberfrax	Al_2O_3-50SiO_2	2.5	2.70	0.098	1.72	250	105	15,230
Nicalon	SiC	10-15	2.55	0.092	2.5-3.0	360-435	180-200	26,100-29,000

Table 4 Selected fiber-reinforced metal-matrix composites

Matrix alloy	Fiber	Vol%	Density		Processing technique
			g/cm^3	**lb/in.^{3}**	
Cu-1-10	MM3000	39	6.13	0.221	Liquid infiltration
	Thornel 50	39	6.36/5.83	0.230/0.211	Liquid infiltration
	Modmor 1	35	6.90	0.250	Liquid infiltration
Cu-Ag-P-Mg	Celion 6000	42	5.94	0.215	Powder metallurgy
	Thornel 300	38	6.19	0.224	Powder metallurgy
Al-4.7Cu (A2010)	Thornel 50	30	. . .	. . .	Liquid infiltration
Al-4.7Cu (A2010)	Thornel 300	30	. . .	. . .	Liquid infiltration
Al-1Mg (6061)	Thornel 50	30	. . .	. . .	Liquid infiltration
Copper	Thornel 50	30	. . .	. . .	Liquid infiltration
Tin	Thornel 50	30	. . .	. . .	Liquid infiltration
Al alloy (ADC 12)	SiC whisker	20-40	. . .	. . .	Squeeze casting
Cu-Zr	PAN type (carbon)[(a)]	3-50	. . .	. . .	Impregnation
Ag-28/20Cu	Thornel 300	42/44	6.58/6.41	0.238/0.232	Liquid infiltration
Ag-28Cu	Celion 6000	40	6.80	0.246	Liquid infiltration
Ag-26Cu-2Ni	Celion 6000	28	7.65	0.276	Liquid infiltration

(a) PAN, polyacrylonitrile

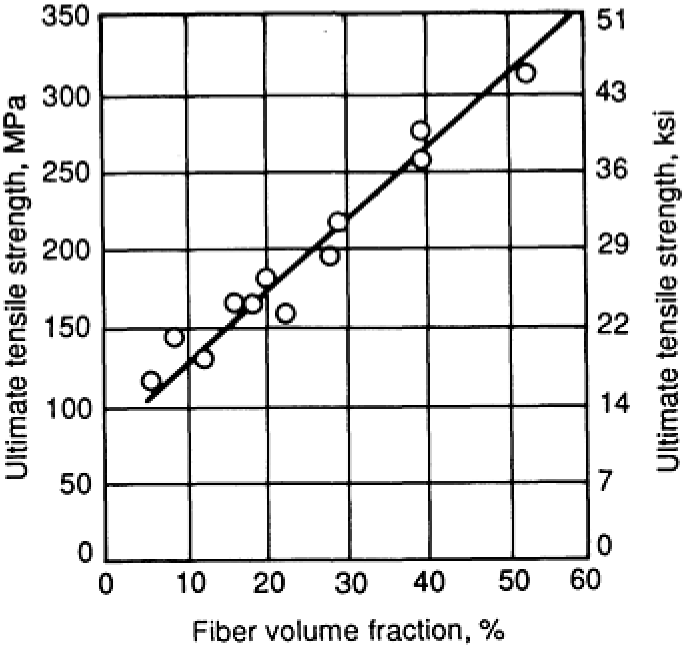


Fig. 3 Ultimate tensile strength of squeeze-cast continuous graphite-fiber-reinforced aluminum alloy composites as a function of fiber volume fraction

The fracture toughness of particulate composites is low, and it decreases with volume fraction of a dispersed phase such as SiC, as shown in Fig. 4 (Ref 10, 23, 28, 29). A weak interface leads to cracks at low strain, and a low fracture

toughness value indicates their easy propagation. In the context of wear, these two properties are critical to the generation and propagation of cracks, resulting in wear debris. Pores in composites may act as preexisting crack nuclei in the system, waiting to become unstable at appropriate stress levels (Ref 30). Porosity thus increases composite wear rate.

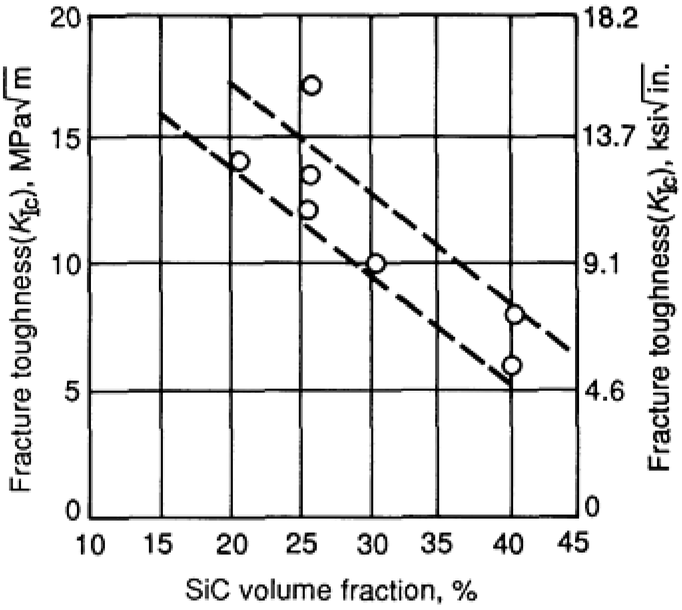


Fig. 4 Fracture toughness of aluminum alloy composites containing SiC as a function of volume fraction of SiC

During tribological interaction between two surfaces, temperature rises significantly at and near the surface, and the mechanical properties of the composites at elevated temperatures may be of direct relevance. The dispersed hard particles are helpful in retaining the high-temperature strength of the matrix (Ref 31). Also, the loss of strength due to dynamic recrystallization is counteracted by the presence of particles. Metal-matrix composites can thus have superior mechanical properties at elevated temperatures, even though the room-temperature properties may not look encouraging in certain systems.

Tribological Behavior of Metal-Matrix Particulate Composites

Friction and Sliding Wear. Friction and wear behavior of metal-matrix composites depends on the nature of particle reinforcements in close relation to the matrix containing them. The particles can be softer or harder compared with the matrix. Ceramic particles generally used as reinforcements include carbon, silicon carbide, and alumina, which have low adhesion to a metallic counterface. The asperity of the counterface can easily plow through softer particles such as char (carbon); it cannot do so through harder particles such as alumina or silicon carbide.

Figure 5 shows the variation of specific wear rate as a function of volume fraction of particles in different systems (Ref 12, 32). Char is a soft particle that contains carbon and other hard mineral oxides, and composites containing char exhibit much wear. Alumina is harder than char but softer than silicon carbide, and an aluminum-alumina composite shows higher wear than an Al-SiC composite but lower wear than an aluminum-char composite. However, aluminum-base composites that contain graphite have the lowest wear, because graphite is not only soft but also shears easily along the basal plane of its hexagonal close-packed lattice in suitable environments and acts as a solid lubricant. Composites containing a solid lubricant such as graphite have low wear because of its transfer onto the tribosurfaces and the subsequent formation of a lubricating film between the matrix of the composite and the counterface. Thus, it may be concluded that the harder the reinforcing particles, the lower the sliding wear in a composite, with the exception of solid lubricating dispersoids of the right crystal structure, which impart low wear rate despite their low hardness. Another interesting observation in Fig. 5 is that wear resistance can be improved by increasing the volume fraction of either hard or soft particles in the range of conditions investigated.

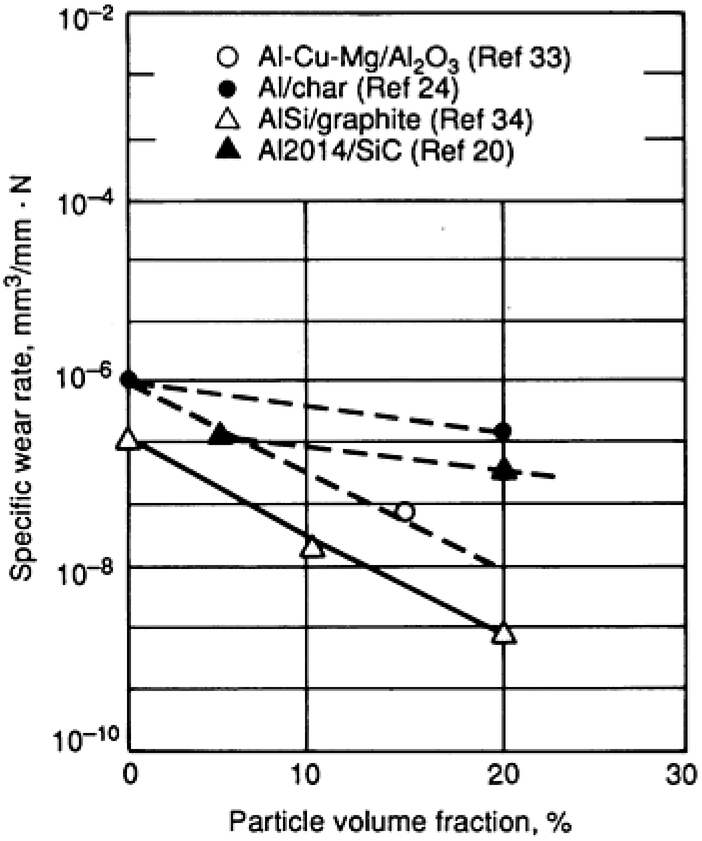


Fig. 5 Specific wear rates for several aluminum alloy particulate composites sliding against steel as a function of particle volume fraction

The friction coefficient values for various composites are given in Fig. 6. It seems that particles with higher hardness result in higher coefficients of friction. Figure 6 can be divided roughly into three overlapping regions (Ref 24, 33, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44):

- *Region 1:* μ= 0.3 to 0.5, the range of values for composites containing hard particles
- *Region 2:* μ= 0.25 to 0.4, the range of values for composites containing particles of intermediate hardness
- *Region 3:* μ= 0.10 to 0.30, the range of values for composites containing soft particles

However, this classification does not reflect matrix characteristics and thus can be applied only to composites of the same or similar matrix. The hard particles in composites result in larger plastic deformation due to asperity-asperity interaction during sliding and contribute to higher coefficients of friction despite reduced contribution of adhesion between the particle and the matrix. In region 3, a lubricating film forms on the mating surface, caused by the smearing of soft solid lubricating particles during the sliding process. In region 2, although a transfer layer may form during sliding, this layer cannot reduce the coefficient of friction significantly because it is not a lubricating film. Solid lubricating films are responsible for lower friction coefficients in composites that contain solid lubricants such as graphite or MoS_2. Rohatgi *et al.* (Ref 45) have published detailed results on the tribological behavior of metal-matrix/graphite-particle composites. Their results indicate that when the volume fraction of graphite in the composite is greater than 20 vol%, the coefficient of friction is close to 0.2 regardless of the matrix. Apparently, above 20 vol% graphite, the graphite film covers the entire surface and prevents direct metal-to-metal contact, as shown in Fig. 7 (Ref 40, 42, 43, 45, 46, 47, 48, 49, 50). Under these conditions, the friction coefficient is not a significant function of the matrix alloy chemistry.

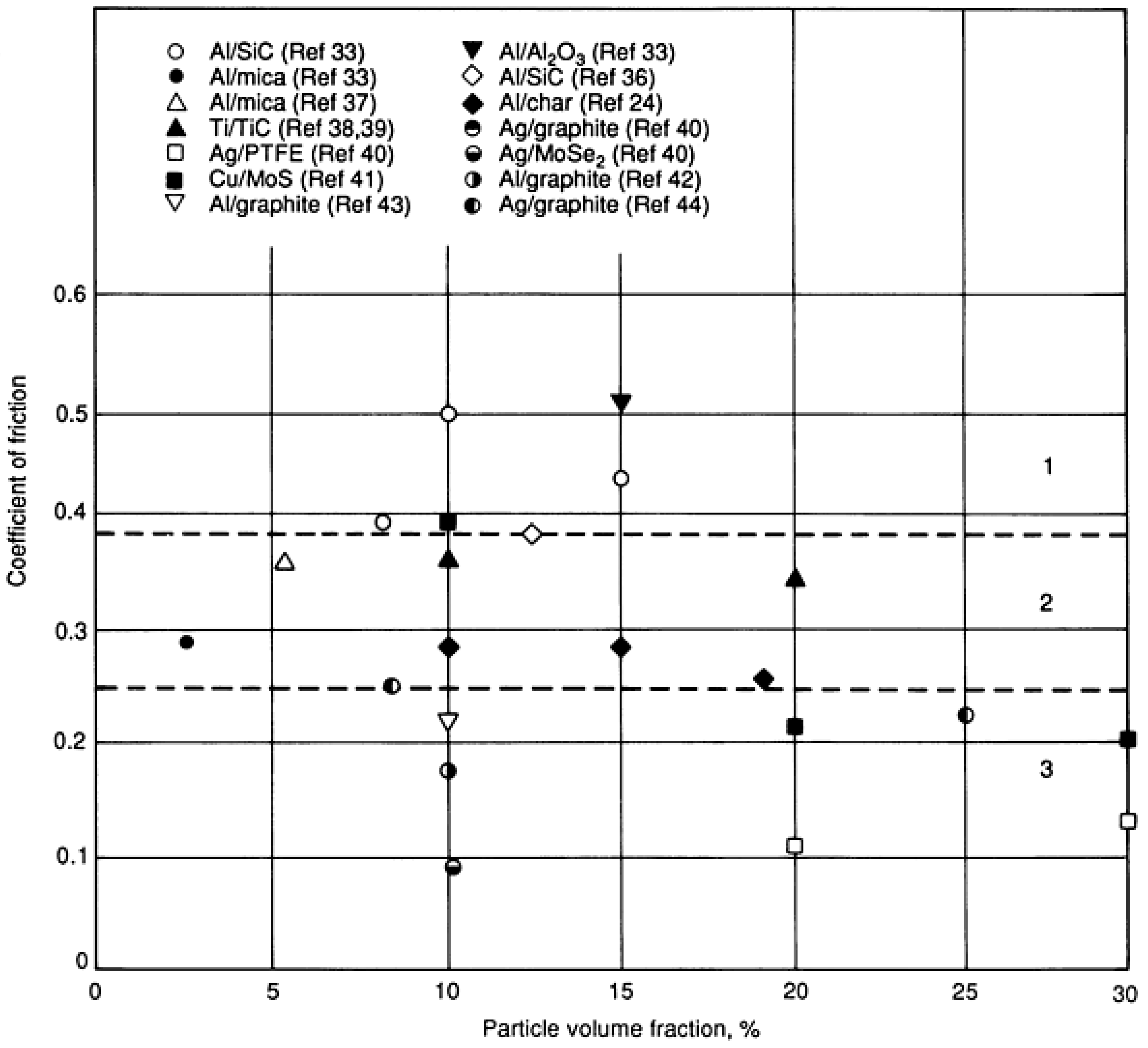
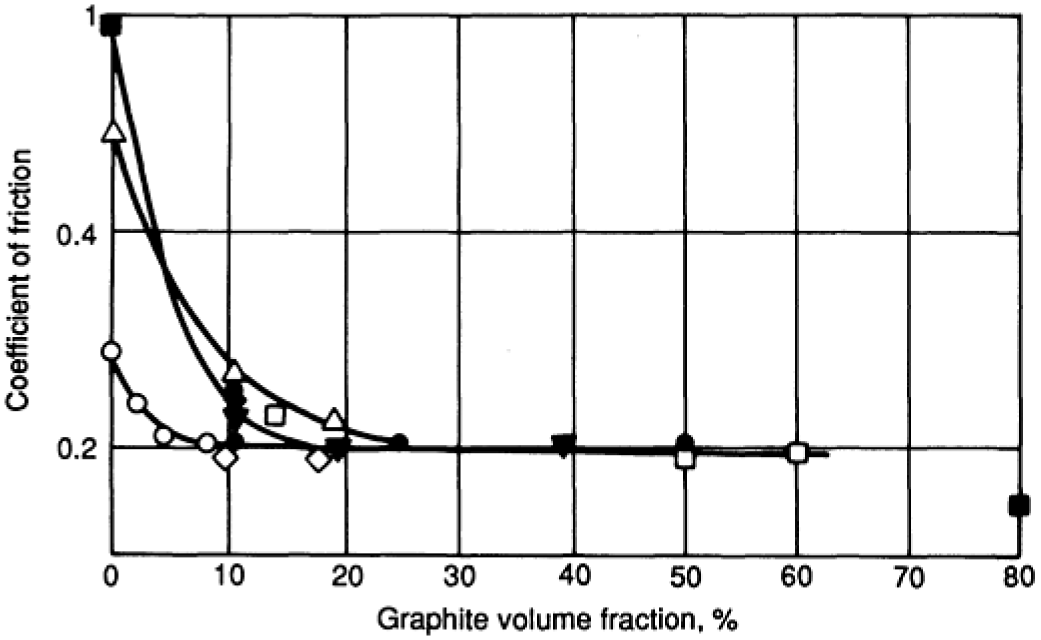


Fig. 6 Coefficient of sliding friction as a function of particle volume fraction in metal-matrix composites sliding against a steel counterface (except those with a silver matrix). See text for discussion of regions 1 to 3.



Symbol	Material	Normal pressure		Sliding speed, m/s	Ref
		MPa	**ksi**		
○	Al-Si	1	0.15	1	43
•	Al-Si	1.6/0.12	0.2/0.017	0.1/0.2	42
△	Al	0.4	0.06	0.5/1	46
▲	Ag	0.3	0.04	13/26	47
□■	Fe-Ni	. . .	. . .	1.2/2.3	40
▽	Cast iron	0.6	0.09	5	48
▼	Cu	0.5/0.05	0.07/0.007	. . .	49
◇	Cu	0.5/2	0.07/0.3	0.6	50

Fig. 7 Coefficient of sliding friction as a function of volume fraction of graphite particles in metal-matrix composites sliding against steel (except those with a silver matrix)

Friction and Abrasive Wear. The abrasion resistance of a material generally is determined by two different types of tests: (1) low-stress tests, such as rubber wheel abrasion tests (RWAT; for details, see ASTM Standards G-65 and B-611) and (2) high-stress tests, such as pin-on-disk and pin-on-drum types of tests. In low-stress tests, the abrading particles do not break; in tests employing high stress levels, the abrading particles break. The abrasive particles are generally rounded sand (SiO_2), crushed quartz, or silicon carbide particles of different grit sizes and shapes. The size and shape of abrading particles, as well as their relative hardness with respect to the composite being tested, are important parameters in determining the extent of wear. Bhattacharrya and Bock (Ref 51) examined the effects of different abrasive particles on the abrasive wear of engineering materials and concluded that alumina is more aggressive than silica sand. In monolithic materials, it has been observed (Ref 52) that wear volume increases rapidly with grit size of abrading particles up to a critical diameter; beyond this size, wear increases at a slower rate. When the size of the abrading particle is less than 1 μm, the wear is no longer by abrasion. In metal-matrix composites, as in monolithic materials, abrasive wear involves gouging, grooving, and plastic deformation caused by penetration of hard abrading particles (Ref 53). The abrading particles may also interact with the dispersed hard particles in the composite. In RWATs, the stress developed during this interaction between hard particles cannot exceed a limiting stress required for local elastic deformation of the rubber wheel to engulf the abrading particle. In high-stress tests employing the rigid backing surface of a disk or drum, the particle interaction can be very severe, leading to widespread fracture of the hard but brittle particles.

Specific abrasive wear rates as a function of volume fraction for several hard particles dispersed in aluminum matrices are shown in Fig. 8 (Ref 14, 16, 37, 54). The wear rates decrease with an increase in volume fraction of particles in a manner

similar to those observed in adhesive wear. For a given volume fraction of particles, composites that contain harder particles exhibit a lower wear rate. Although the values of abrasive wear show considerable scatter, the specific wear rate for aluminum-base composites containing hard particles is 10^{-6} to 10^{-7} $mm^3/mm \cdot N$ when the volume fraction of particles is less than 40%. Tawarada *et al.* (Ref 55) have pointed out that the specific wear rate for Al-4.5Mg alloy containing more than 50% Al_2O_3 particles is extremely low, of the order of 10^{-9} to 10^{-10} $mm^3/mm \cdot N$.

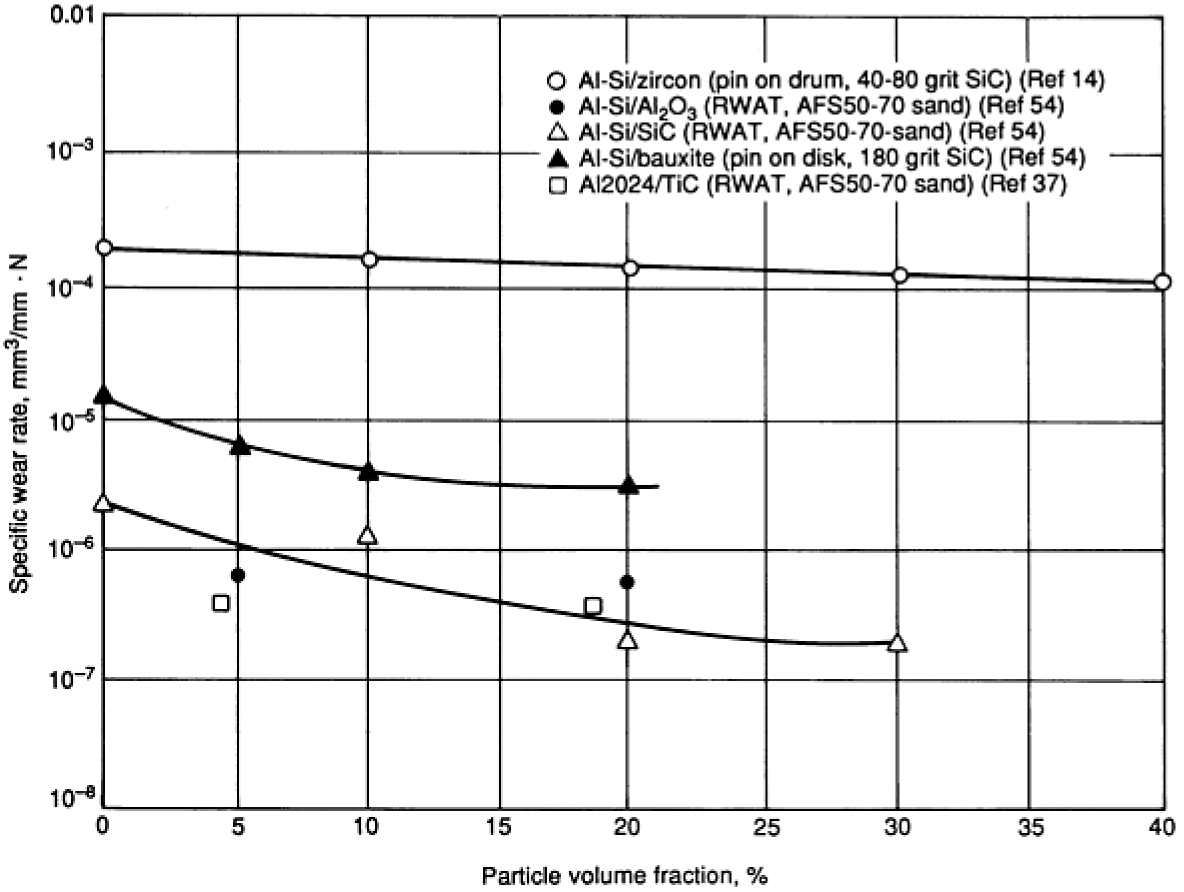


Fig. 8 Specific abrasive wear rate as a function of particle volume fraction in aluminum alloy composites

The improvement in abrasion wear rate of various metal-matrix composites over that of the base alloy is a function of the volume of different reinforcing particles when subjected to RWATs (Fig. 9) (Ref 56). The properties of the dispersoid did not significantly affect these results, probably because of low-stress interaction between the dispersoid and the abrading particles at a level much lower than the hardness of the dispersoids. Under high-stress conditions, the wear rate decreases almost linearly with the volume fraction of dispersoids in a composite. Similar results have been reported for copper-base hard-particle composites (Ref 57). Apparently, the particle spacing decreases with an increase in volume fraction, and any cracks formed easily propagate through the reduced distance in the ductile matrix region to reach the neighboring particles; this results in faster wear of the composites (Ref 58, 59, 60).

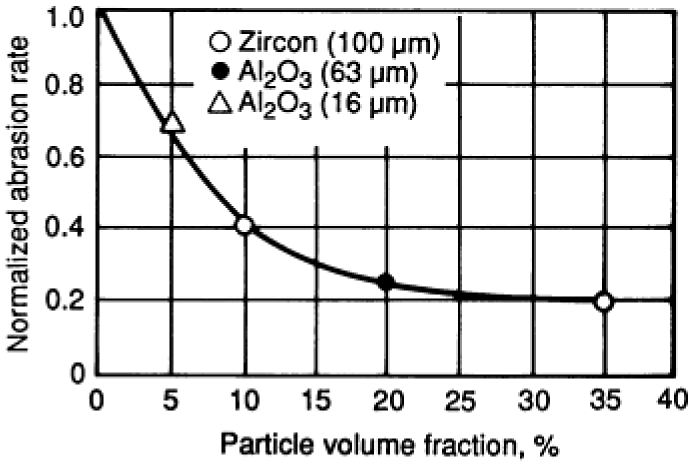
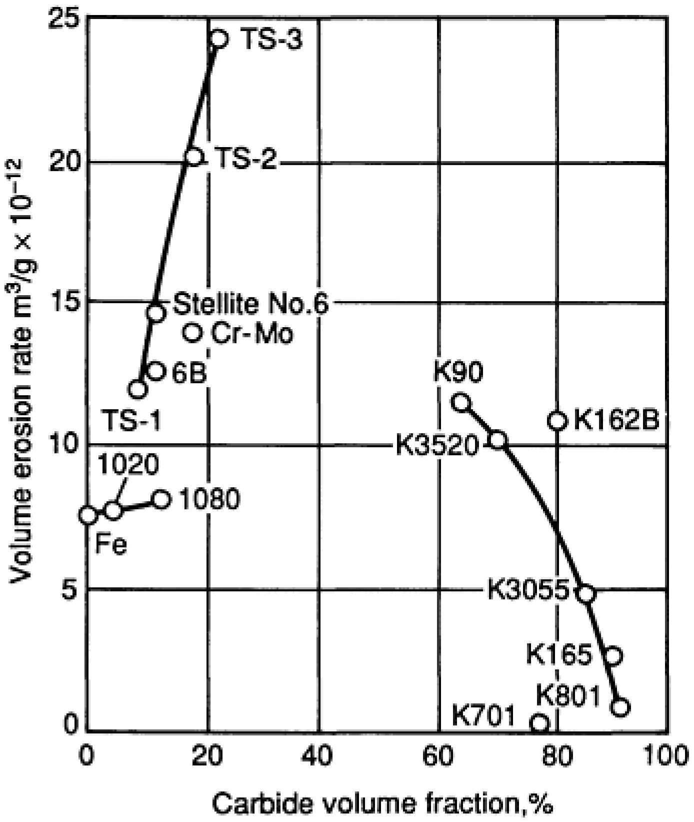


Fig. 9 Normalized abrasion rate as a function of particle volume fraction for aluminum alloy composites under low-stress abrasive wear

Very few publications on steel-base particulate composites are available in open literature. Recently, Halley *et al.* (Ref 61) have reported that the abrasive wear volume of tool steel (0.6-0.9C-10Cr-Mo) with 17 vol% TiC particle composites was lower by 54% compared with similar tool steels. Other investigators (Ref 62) also report that wear rate decreases with an increase of TiC volume fraction for steel-base TiC particle composites.

Under abrasive wear, the coefficient of friction in composites containing hard particles is also a function of the volume fraction of dispersoids. As the volume fraction of hard particles increases, the coefficient of friction decreases for both aluminum-base (Ref 42) and copper-base (Ref 57) composites. The increase in the volume fraction of hard particles in the matrix alloy reduces the area fraction of matrix, and thus there is enhanced ceramic-ceramic contact. The friction coefficient for ceramic-ceramic contact can be 0.25 or less (Ref 57, 63), and that for metal-metal contact is generally much higher (Ref 57, 64). The increase in the relative contact area involving ceramic particles increases the contribution of ceramic-ceramic contacts, resulting in an overall decrease in friction coefficient.

Friction and Erosive Wear. In most cases, the abrasive wear resistance of composites increases with increasing particle volume fraction of hard particles, as discussed above. However, in the case of erosive wear (for details, see ASTM Standard G-76), opposite results have been reported (Ref 65, 66, 67, 68), as shown in Fig. 10. The steady-state erosion rate for 75 to 200 μmquartz erodent at 60 ms^{-1} and an impact angle of 60° increases with an increase of carbide content in both iron- and cobalt-base composites. Similar results have been obtained for other impact angles. In addition, if the matrices are changed from metals to ceramics, the erosion rate decreases with the content of carbides in ceramics (Fig. 10). Ball and Patterson (Ref 67) have pointed out that the minimum in erosive wear rate appeared at 34 vol% binder in WC-Co cermets. A similar result has been reported in the same system for slurry erosive wear (Ref 68). It has been established that, in erosive wear, the angle of impact is very important. In ductile materials, erosive wear is maximum at an impact angle of 15° to 30° for brittle materials, however, this angle is about 90°. Goretta *et al.* (Ref 69) have found that erosive wear rates reach a maximum when the impact angle of the erodent Al_2O_3 was 15° to 30° in Al-20SiC particle composites, and the wear rates of the Al-SiC composites are higher than that of the base alloy under impact angles from 15° to 90°. It is believed (Ref 65, 66, 67, 68, 69, 70) that erosion rates are affected by particle size, particle agglomeration, heat treatment, erodent hardness, impact angle and velocity, and volume fraction of hard particles (or second phase). It has been claimed that erosion rates of metallic alloys are not likely to be reduced by additions of hard particles, except for cases when relatively soft particles (such as quartz abrasive) are used as erodents (Ref 66).



Alloy	Composition	Carbide, vol%
Composites		
Pure iron	Fe	0
Spheroidized 1020	Fe-0.2C	3.1
Spheroidized 1080	Fe-0.8C	12.1
Stellite No. 6	Co-30Cr-4W-1.1C	10.4
Haynes 6B	Co-30Cr-4W-1.1C	10.4
Tristelle TS-1	Fe-30Cr-10Ni-12Co-5Si-1C	8.3
Tristelle TS-2	Fe-35Cr-10Ni-12Co-5Si-2C	16.8
Tristelle TS-3	Fe-35Cr-10Ni-12Co-5Si-3C	21.4
Cr-Mo (white iron)	Fe-20Cr-2.5Mo-2.6C	23.1
Cermets		
K90	WC-25Co	63.1
K3520	WC-20Co	69.5
K3055	WC-10Co	83.7
K701	WC-10Co-4Cr	76.2
K801	WC-5.7Ni-0.4Co	89.9
K162B	TiC-6NbC-2WC-25Ni-7Mo	78.5
K165	TiC-7WC-2.5NbC-9Ni-9Mo	88.8

Fig. 10 Erosion rate as a function of particle volume fraction for iron- and cobalt-base composites and cermets

Mechanisms of erosion include fracture-toughness-dominated wear and hardness-dominated wear. The former is applicable for ceramic composites with a lower amount of metal binder, but the latter is the criterion suitable for erosion wear in metal-matrix composites.

Effect of Particle Size. Figure 11 shows the effect of particle size on specific wear rates of composites under sliding wear conditions. The larger the particles, the lower the wear rate for composites that contain either hard or soft particles.

Similar results have been reported by Suwa *et al.* (Ref 21) for aluminum-graphite systems. It should be pointed out that for small sizes of hard particles ($\leqslant$20 μm), a change in particle size may not significantly affect wear results.

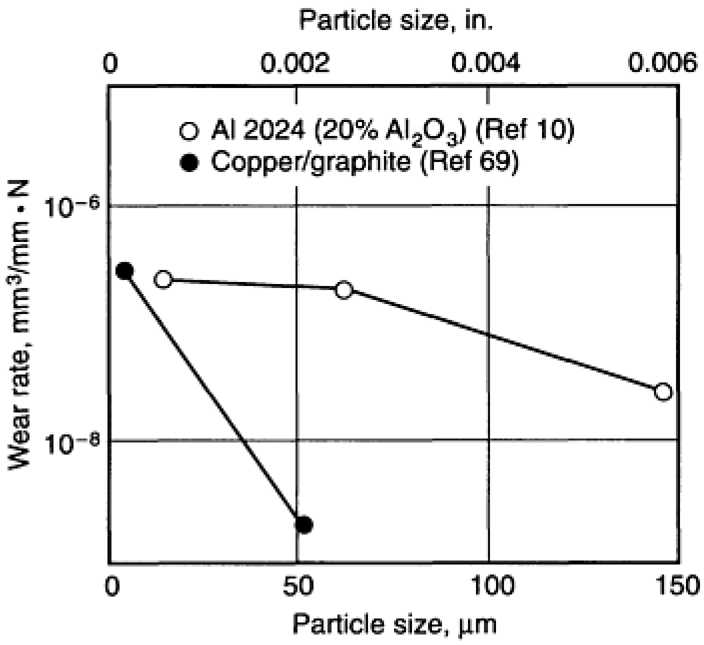


Fig. 11 Specific wear rate as a function of particle size for metal-matrix composites under sliding wear

The coefficient of friction for composites containing soft particles is affected by particle size. Tsuya (Ref 41) has reported that the value of the friction coefficient in sliding decreased from 0.4 to 0.2 in Cu-20WS_2 composites when the size of WS_2 increased from 0.1 to 0.2 μm. The effect of particle size on sliding friction coefficient is not clear for composites containing hard particles (Ref 20).

Desai *et al.* (Ref 71) have indicated that, under abrasive wear, dispersions of coarser carbide particles (chromium-enriched M_7C_3 carbide) reduced the wear of cobalt-base composites, as shown in Fig. 12. Similar results have been reported by Bhansali and Mehrabian (Ref 54) and by Tawarada *et al.* (Ref 55) for aluminum-base composites.

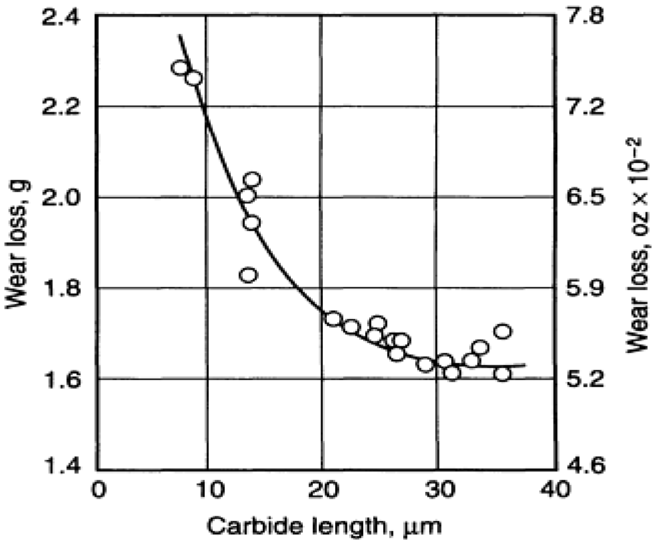


Fig. 12 Wear loss as a function of particle size (M_7C_3) for cobalt-base composites under the rubber wheel

The effect of hard-particle size on friction and abrasive wear in aluminum-base composites has also been studied (Ref 36). In aluminum-magnesium alloy base composites that contain SiC, the friction coefficient decreased as the average particle size decreased for a constant percentage of SiC particles.

Effect of Testing Parameters. The effect of normal load on wear of composites containing hard particles and particles of intermediate hardness has been reported for both adhesive (Ref 20, 24, 25, 38) and abrasive wear (Ref 14, 54, 55). Both wear rates increase as the normal load increases (Fig. 13). Similar results have been reported by Nath *et al.* (Ref 38), Murali *et al.* (Ref 24), and Yang and Chung (Ref 16). However, for certain composites that contain a high volume fraction of soft particles such as MoS_2 and WS_2, the wear rate decreases up to a limiting value of contact pressure and then increases, as shown in Fig. 14. This result is quite similar to that obtained during metal-to-metal sliding using solid lubricants (Ref 70), indicating that a lubricating film may form at the mating surfaces during sliding wear of composites that contain soft sulfide particles.

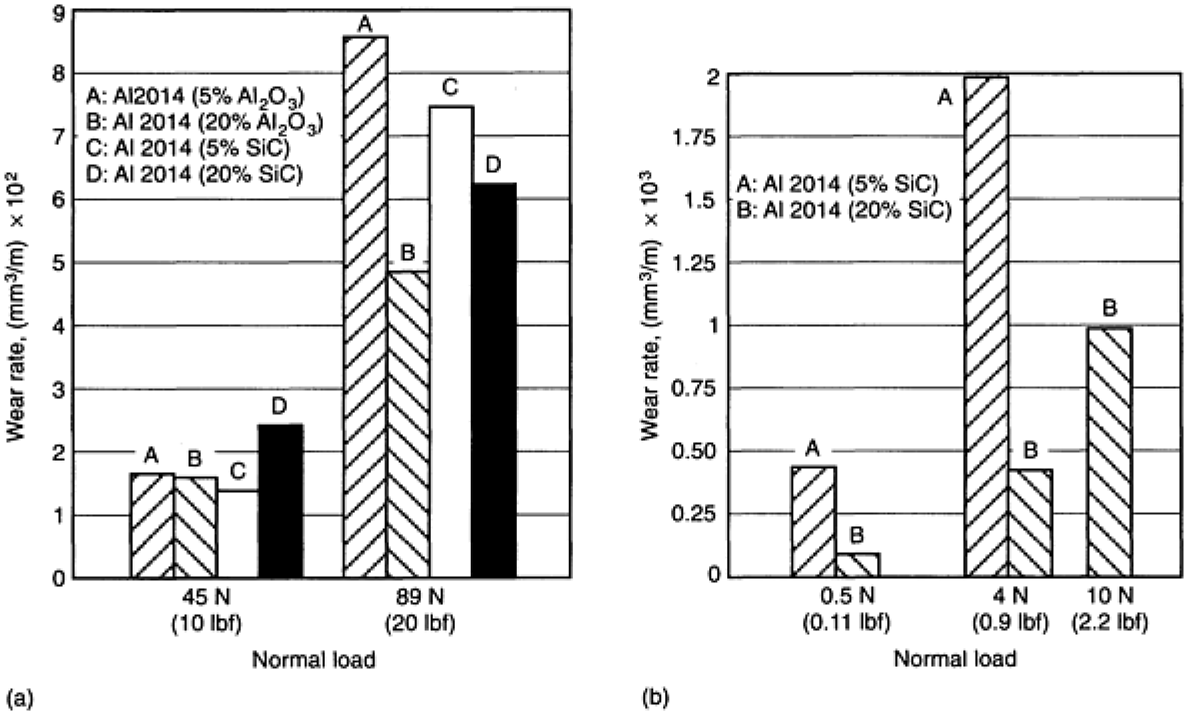


Fig. 13 Wear rate as a function of normal load in metal-matrix composites. (a) Under adhesive wear against steel. (b) Under abrasive wear in dry sand RWAT

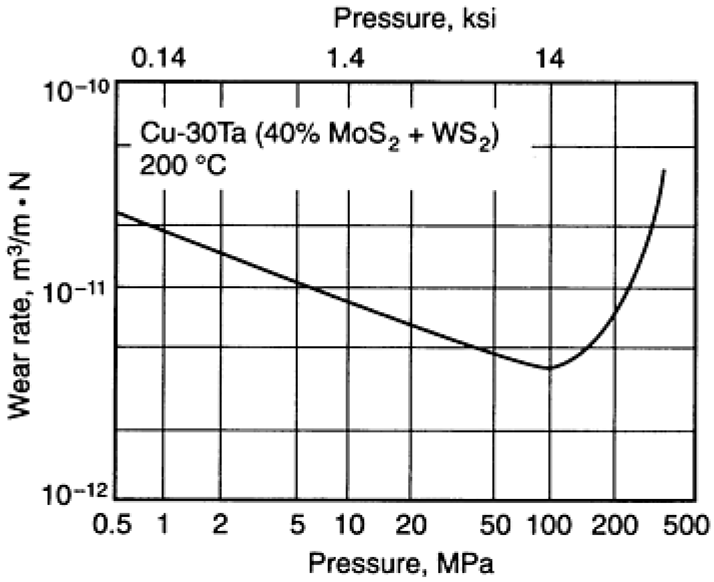


Fig. 14 Specific wear rate as a function of normal contact pressure for Cu-30T a base soft-particle composites under sliding wear

In composites that contain a lower volume fraction of soft particles, the tribosurface may not be completely covered by a film, and the wear volume increases with normal pressure, as shown in Fig. 15 (Ref 45). Similar results have been obtained for aluminum-lead composites by Mohan *et al.* (Ref 72).

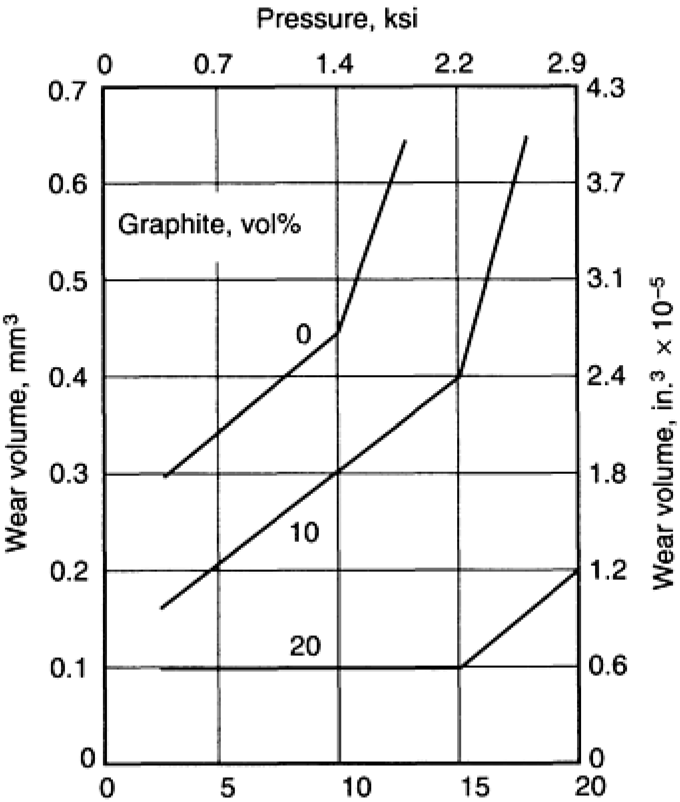


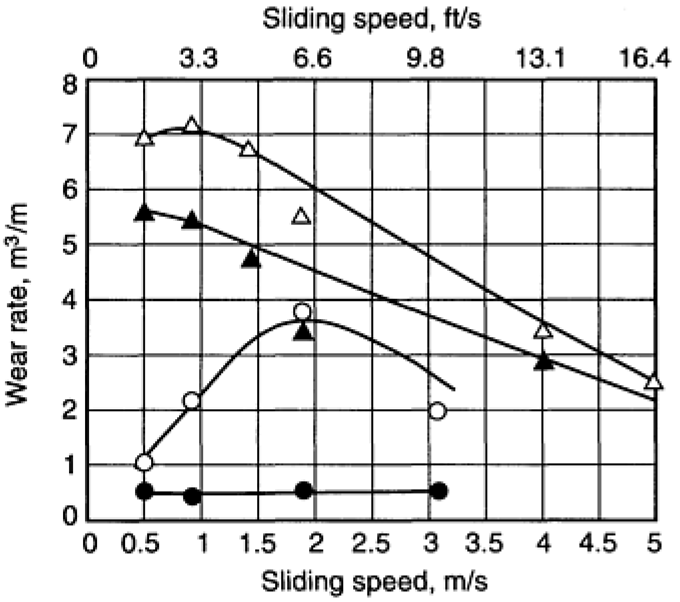
Fig. 15 Variation of wear with normal pressure in copper-base composites containing different amounts of

graphite particles.

The effect of normal load on friction under sliding wear for composites containing hard particles has been reported (Ref 16). The coefficient of friction tends to decrease as the normal load increases.

Keshavaram and co-workers (Ref 35) have reported that in aluminum alloy composites containing 10% glass, the wear rate decreases with an increase in sliding velocity (0.1 to 3 m/s) under adhesive wear. On the other hand, Suzuki *et al.* (Ref 73) have reported that, under abrasive wear, the wear (in microns) of an Al-12Si-50Al_2O_3 composite was almost constant as the sliding velocity increased from 3 to 8 m/s. When these results are converted into wear rate, the same trend observed by Keshavaram *et al.* (Ref 35) is evident.

The effect of sliding velocity on wear rate is more complex for composites that contain soft particles. Rohatgi *et al.* (Ref 45) have reported results for aluminum-graphite systems. The wear rate for an aluminum alloy composite containing 5% graphite decreases with increasing sliding speed (0.5 to 5 m/s) (Ref 74). Sliding speed does not affect wear rate significantly if the volume fraction of graphite particles increases to 15% (Ref 46), as shown in Fig. 16, indicating that the sliding surface is largely covered by a graphite film.



Base alloy/ composite	Material	Graphite content, %	Pressure		Ref
			MPa	ksi	
○/•	Al-12Si	15	0.4	0.06	46

Fig. 16 Wear rate as a function of sliding speed for aluminum alloy composites containing graphite particles

Effect of Contacting Conditions. In general, composites that contain soft particles are tested only under sliding wear conditions, and composites that contain hard particles are tested under various conditions, including sliding, rolling and abrasive wear.

Sliding wear conditions are based on differences in the nature of relative movement between surfaces: reciprocating and rotational. Abrasive wear can be divided into two-body abrasive wear, such as rotation against a ceramic disk or ceramic particle coatings, and three-body abrasive wear, such as RWAT with ceramic particles fed between the specimen and the rotating rubber wheel.

For the same material, different contacting conditions will affect the wear result and even the wear mechanisms (Ref 75). Pan *et al.* (Ref 12) have reported that for aluminum alloy composites containing 15% ceramic particles, the specific wear rate under rolling contact is higher than that under sliding contact, as shown in Table 5. The wear rate under abrasive wear for the same matrix alloy is also included in Table 5 for reference. In rolling contact, the roller is subjected to a cyclic stress, and thus fatigue failure appears to be the dominant wear mechanism. The material is removed by debonding and the removal of ceramic particles, and by the formation of fatigue cracks at pitting spots on the rolling surface (Ref 12).

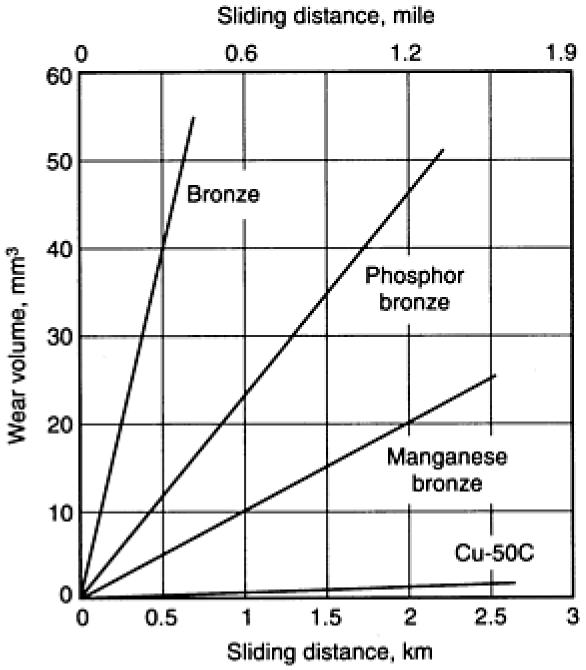
Table 5 Effect of contact conditions on wear rate in aluminum alloy hard-particle composites

Composite		Normal load, N	Sliding speed, m/s	Specific wear, $mm^3/m \cdot N$
Matrix	**Vol%/ particle**			
Sliding contact[a]				
Al 2014	15/Al_2O_3	168	2	1.3×10^{-9}
Al 356	15/SiC	168	2	1.2×10^{-9}
Rolling contact[a]				
Al 2014	15/Al_2O_3	1112	3.12	3.5×10^{-8}
Al 356	15/SiC	1112	3.12	7.6×10^{-8}
Abrasive contact				
Al 2014	20/Al_2O_3	134	Flow rate of silica, 280 g/min	9.4×10^{-4}
Al 2014	20/SiC	134	Flow rate of silica, 280 g/min	7.4×10^{-4}

(a) Tested against a steel counterface

Tribological Behavior of Fiber-Reinforced Metal-Matrix Composites

Adhesive Wear Conditions. Figure 17 shows the wear volume for dry reciprocating sliding wear, as a function of sliding distance, in different copper alloys and copper-carbon continuous-fiber composites, with the fibers aligned perpendicular to the sliding surface. Compared with the copper alloys, the copper composite containing 50% reinforced carbon fibers exhibits a significantly lower wear rate (Ref 76).



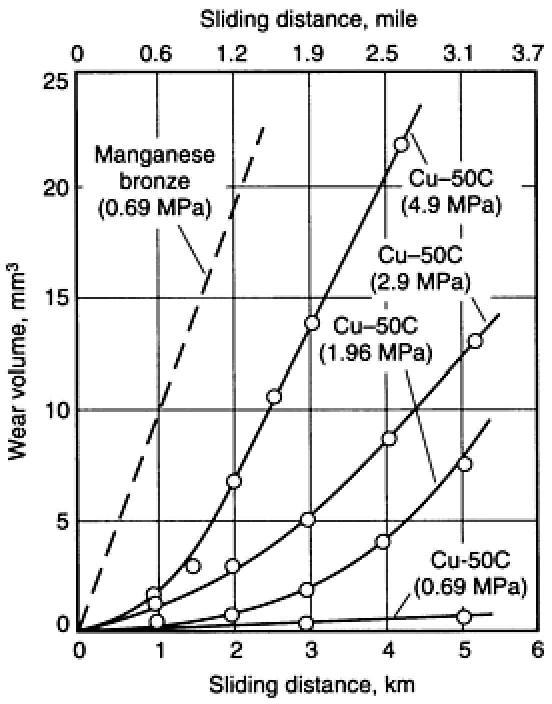


Fig. 17 Wear behavior of copper alloys and copper-carbon fiber composites. (a) *P* = 0.69 MPa (0.1 ksi). (b)

0.69 MPa ⩽P ⩽4.9 MPa (0.1 ⩽P ⩽0.7 ksi)

The effect of volume fraction of carbon fibers on the friction and wear behavior of copper alloy composites is shown in Fig. 18. The friction coefficient and wear rate decrease when the volume fraction increases from 30 to 40%. Similar wear results have been reported for aluminum alloy composites containing SiC whiskers (Fig. 19). Wear is reduced as the volume fraction of SiC whiskers increases from 0 to 20%; up to 40 vol% SiC, further decrease in wear is small. Although there is a wide scatter in the wear results for fiber-reinforced metal-matrix composites, wear resistance of metals or alloys can be improved by the introduction of a proper amount of fibers (Ref 76, 77, 78, 79, 80, 81, 82, 83). The effect of volume fraction of fibers on the coefficient of friction is less clear. It appears that the coefficient of friction tends to decrease with increase in volume fraction of graphite in copper-graphite composites (Fig. 18), but friction increases with whisker content in aluminum alloy composites that contain SiC whiskers (Fig. 19). Testing parameters also affect the wear rate of fiber-reinforced metal-matrix composites. Wear rate increases with an increase in normal load (Ref 76, 82, 83), and there is a transition from mild to severe wear at a critical load, as observed in Al-20SiC whisker composites (Ref 78). A similar transition has been reported in copper-carbon fiber composites, depending on the volume fraction of fibers (Ref 76).

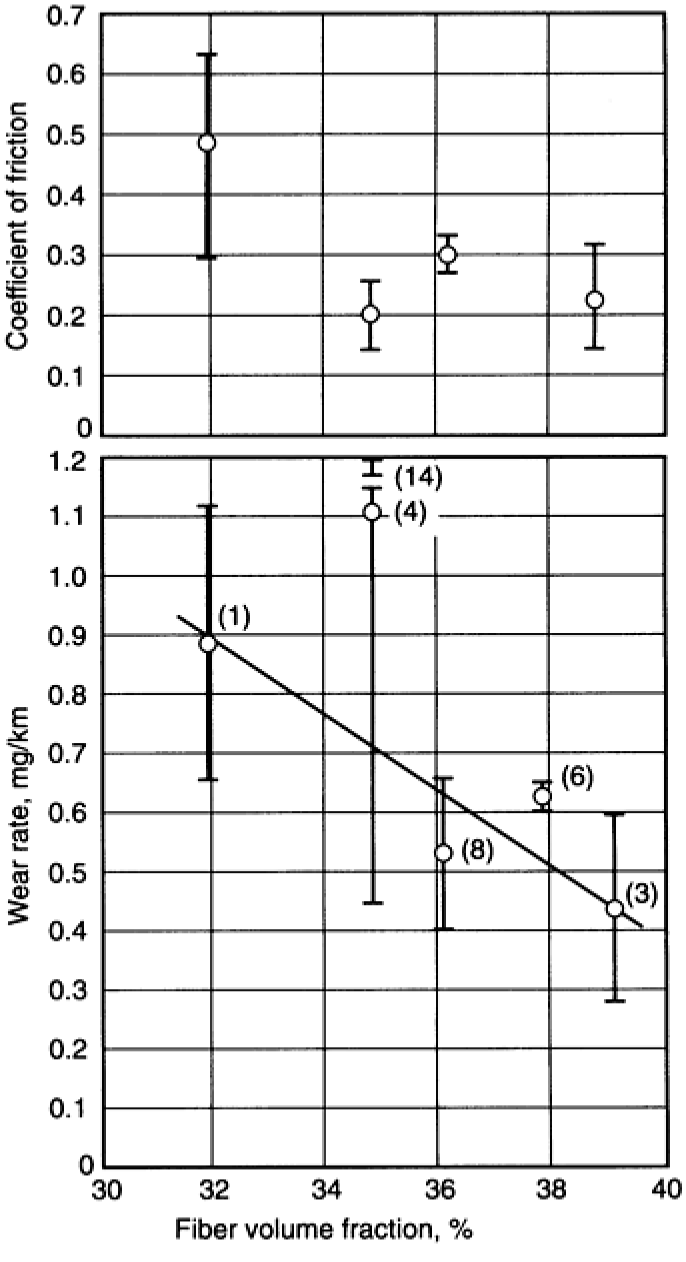


Fig. 18 Effect of fiber fraction on wear and friction of Hm3000/bronze carbon fiber. Numbers in parentheses

indicate tin content of bronze matrix.

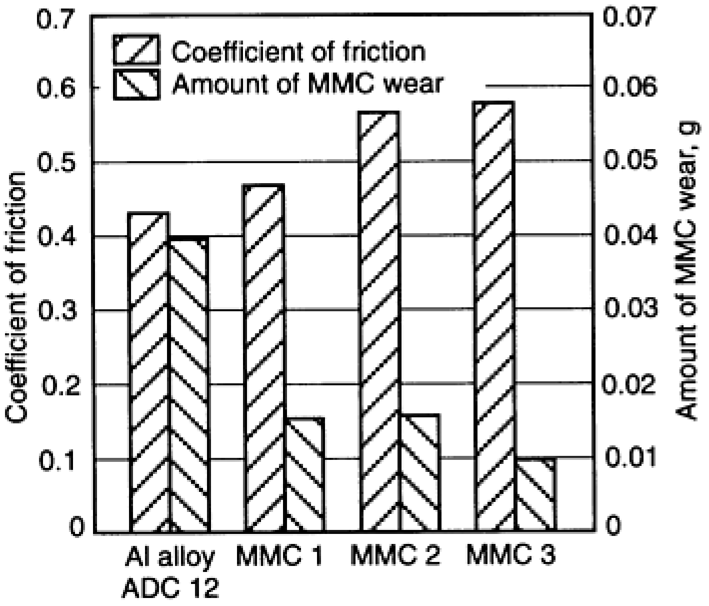


Fig. 19 Friction and wear characteristics of metal-matrix composites (MMC) on steel. The composites all had an ADC 12 matrix and were reinforced with the following volume fractions of SiC whiskers: 20% (MMC1), 30% (MMC2); 40% (MMC3). Test conditions: sliding velocity, 2 m/s; pressure, 0.1 MPa (0.015 ksi); duration, 1 h

Abrasive Wear Conditions. Recently, Verma and Dorcic (Ref 84) reported that the abrasive wear resistance of aluminum alloy A206 with discontinuous fibers of Al_2O_3 and SiO_2 is comparable to induction-hardened AISI 1345 steel (Table 6). The tests in Table 6 were performed using a dry sand/rubber wheel abrasion machine (ASTM Standard G65-B). Further studies are needed to determine general trends in the abrasive wear behavior of fiber-reinforced composites.

Table 6 Comparison of wear rates of AISI 1345 and squeeze-cast composites

Material	Specimen no.	Initial weight, g	Final weight, g	Weight loss, g	Average weight loss, g	Average volume loss, mm
AISI 1345	1	114.4135	114.1204	0.2931	0.2779	35.4
	2	114.3513	114.0807	0.2706		
	3	114.3439	114.0739	0.2700		
A206-T7 + ceramic foam	1	43.7071	43.0129	0.6942	0.6885	242.00
	2	48.0881	47.4054	0.6827		

Effect of Fiber Type and Orientation. Graphite fibers from different precursors affect the friction and wear behavior of fiber-reinforced metal-matrix composites (Ref 80, 83), as shown in Table 7. The effect of fiber orientation on friction and wear is shown in Fig. 20. For copper alloy composites containing 50% carbon fiber, wear volume increases with sliding distance for any fiber orientation. However, the wear volume in composites with randomly arranged fibers is extremely low (Fig. 20), and the wear volume of composites with fibers vertical to the sliding surface is lower than that of composites with fibers parallel to the sliding surface. This finding is supported by results obtained by Eliezer and co-workers (Ref 80). Amateau *et al.* (Ref 83) have shown that the orientation of fiber did not significantly affect the wear rate and coefficient of friction of aluminum alloy composites containing graphite fibers. Although experimental results appear to be contradictory for different systems, it is generally believed that fiber orientation affects wear results, not only for metal-matrix composites but also for polymer-matrix composites. It appears that lower wear rates and coefficients of friction are obtained in composites with fibers arranged randomly or normal to the sliding surface; for composites with fibers parallel to the sliding surface, however, friction and wear also depend on fiber orientation relative to sliding direction.

Table 7 Effect of graphite fiber precursor on friction and wear in aluminum alloy composites

Composite		Precursor	Wear rate[(a)], 10^{-14} m^3/s	Coefficient of friction
Matrix	**Fiber (normal to sliding surface)**			
A201	Thornel 50	Rayon	43.1	0.28
	Thornel 300	PAN	7.1	0.48
A16061	Thornel 50	Rayon	36.2	0.25
	Thornel 75	Rayon	32.9	0.25
	Hercules A	PAN	3.1	0.52

(a) Tests were conducted at a load of 6 N, a linear sliding speed of 0.684 m/s, and a sliding time of 6 min.

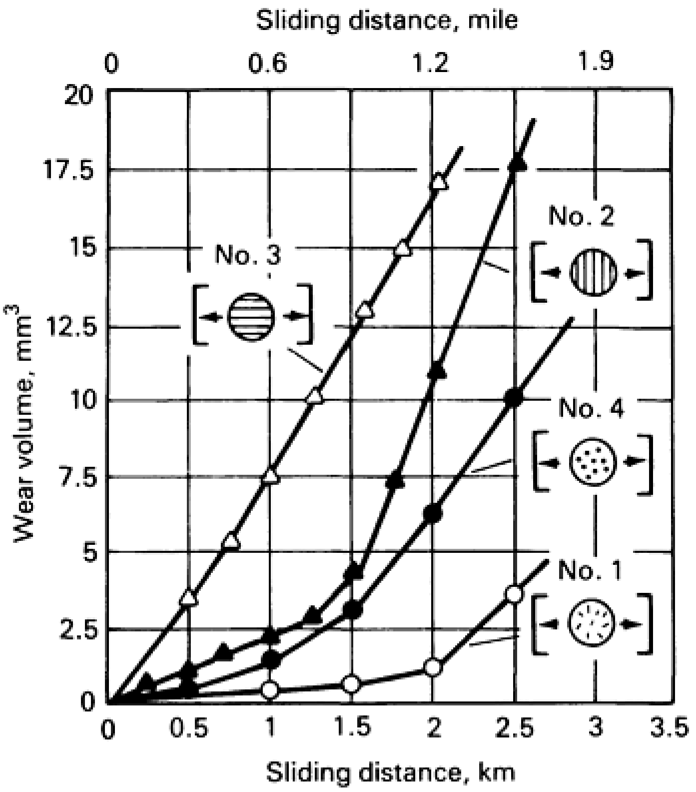


Fig. 20 Effect of sliding direction and fiber arrangement on wear volume of copper-carbon composite. *P* = 4.9 MPa (0.7 ksi)

Kuniya *et al.* (Ref 76) have explained the effect of fiber orientation in copper-base composites. They presumed that because of weak bonding between the fibers and the matrix, the fibers easily debonded when parallel to the sliding direction. Arikan and Murphy (Ref 85) have observed similar behavior in metal-matrix composites with randomly dispersed fibers. Fibers parallel to the wear surface are easily debonded and removed during wear testing, but fibers perpendicular to the wear surface break off in small sections.

Wear Mechanisms

Mechanisms of Sliding Wear in Metal-Matrix Composites Containing Hard Particles. You and co-workers (Ref 86) have reported three different stages of wear in aluminum alloy A2024 base composites that contain 20% SiC particles, when sliding against a steel counterface. Scanning electron microscopy (SEM) reveals these three stages: (1) formation of steel filings (steel debris) on the testing pin, caused by microcutting of the counterface by SiC particles; (2)

packing of the filings on the sliding pin surface between particles; and (3) gross pin surface deformation with loss of SiC particles. The onset of steady-state wear is associated with the end of the second stage or the beginning of the third stage. A transfer layer--a mixture of iron debris, SiC fragments, and aluminum matrix debris--forms on the sliding surface of the pin under steady-state wear. A high level of strain in the subsurface region with a high density of dislocations has been observed up to a depth of 150 μm. The formation of a transfer layer on the sliding surface has been reported by Keshavaram *et al.* (Ref 35) in aluminum alloy composites containing glass particles, by Murali *et al.* (Ref 24) in aluminum alloy composites containing char, and by Nath and Rohatgi (Ref 17) in aluminum alloy composites containing mica particles. Observations by SEM suggest that the debonding of particles is not a dominant mechanism of sliding wear (Ref 16, 20, 85). However, the formation of a transfer layer results in grinding of the sliding surface by a mixture of this layer.

A wear model has been proposed that is based on an adhesive wear mechanism (Ref 76, 78, 80) and on the assumption that each phase of a composite (matrix and particle) wears out independently and contributes to the overall wear rate (Ref 56):

$$K_c = K_m f_m + K_p f_p \quad \textbf{(Eq 1)}$$

where K is the wear coefficient, f is the volume fraction, and the subscripts c, m, and p represent composite, matrix, and particle, respectively. Because the value of K_p is likely to be several orders of magnitude lower than K_m under experimental conditions, Eq 1 can be simplified to:

$$K_c = K_m f_m = K_m(1 - f_p) \quad \textbf{(Eq 2)}$$

This model does not explain the effect of particle hardness, which has been identified as a very influential factor in wear, and should be modified to include the active role of particles in reducing wear, as evident in Fig. 5 for composites containing SiC and Al_2O_3. Considering the hardness of particles, a modified equation has been proposed:

$$K_c = -\alpha f_p + K_m \quad \textbf{(Eq 3)}$$

where α is the coefficient factor and can be obtained from the slope of the wear curves.

Mechanisms of Abrasive Wear in Composites Containing Hard Particles. Abrasive wear usually can be classified on the basis of the nature of interactions, such as two body or three body. For two-body abrasive wear, Saka and Karalekas (Ref 57) have shown that a delamination model is feasible for copper-base hard-particle composites, as shown in Fig. 21. In this model, wear resistance is strongly dependent on fracture toughness and not merely on composite strength, because crack propagation is the controlling factor (Ref 58, 59, 60).

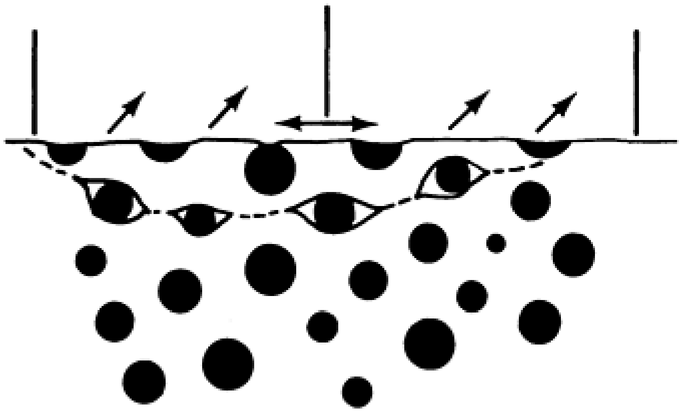


Fig. 21 Schematic diagram showing the delamination process in abrasive wear of metal-matrix composites containing hard particles

In three-body abrasive wear, abrasive particles such as silica sand tend to break down and bury themselves in alloys with relatively soft matrices. Very fine silica particles trapped below scratches have been observed under SEM (Ref 51). However, hard particles in composites resist scratching because of their high hardness under low loads and crack into small pieces under heavy loads. The broken particles remain embedded in the matrix during wear. In this case, composite strength and fracture toughness both play important roles. Similar results have been observed by Yang and Chung (Ref 16) for aluminum alloy composites that contain bauxite particles. Some investigators report that micromachining is still the dominant mechanism of material removal during abrasion wear in composites. Removed chips in the debris are approximately the same size as the hard particles.

Mechanisms of Sliding Wear in Composites Containing Soft Particles. A typical composite in this group is the metal-matrix/graphite-particle composite. The wear mechanisms in composites containing graphite particles are discussed in detail in Ref 45. Similar systems include composites containing dispersoids such as MoS_2 and PTFE. Formation of a lubricating film on the tribosurface is considered to be responsible for the low friction and wear of composites belonging to this group. The area covered by a lubricating film, which is related to the volume fraction of soft particles, is the controlling factor. Tsuya (Ref 41) has measured the concentration of MoS_2 in the solid lubricating film formed during sliding wear in the Cu-MoS_2 system as a function of the volume fraction of MoS_2 in the composite (Fig. 22). More than 60% of the sliding surfaces is covered by MoS_2 during adhesion wear testing in Cu-20MoS_2 composites, and more than 40% of the sliding surfaces is covered by MoS_2 in Cu-10MoS_2 composites. These results are in general agreement with Auger electron spectroscopic observations of aluminum alloy composites containing 10% graphite (Ref 42).

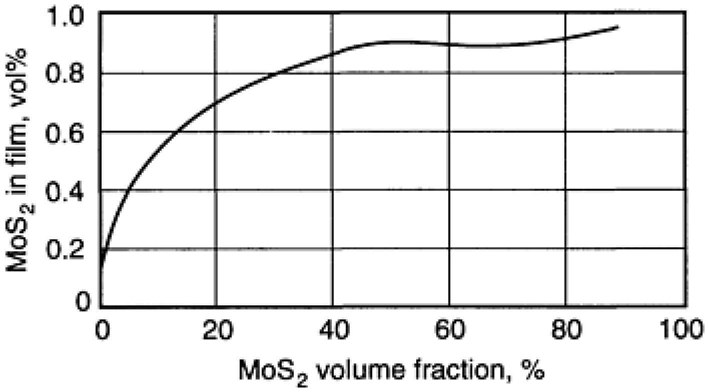


Fig. 22 Concentration of MoS_2 in the surface film formed during sliding as function of MoS_2 volume fraction for copper-base composites. Source: Ref 41

Conclusions

Metal-matrix particle composites have been classified into two groups: those containing hard particles with hardnesses in the range of 4 to 30 GPa (580 to 4350 ksi) and those containing soft particles with hardnesses less than 2 GPa (290 ksi). The tribological properties and relevant mechanical properties which influence wear and friction have been summarized and discussed for these two groups. The main conclusions that can be reached include the following:

- The ultimate tensile strength (UTS) of metal-matrix composites increases with an increase in volume fraction of hard particles, and decreases with an increase in volume fraction of soft particles. The fracture toughness and ductility of composites decrease with the increase in the volume fraction of both hard and soft particles
- Under sliding wear against a steel counterface, the wear rate of metal-matrix composites decreases with an increase in volume fraction of either hard or soft particles if the bonding between the matrix and the particles is adequate
- The coefficient of friction of the composites under sliding wear can be roughly divided into three ranges: In range I, the coefficient of friction of the composites containing hard particles is between 0.3

and 0.5 and is close to that of the base alloys; in range III, the coefficient of friction of composites containing soft particles is between 0.1 and 0.3 and is similar to the coefficient of friction of the soft particles if the volume fraction of soft particles is over 20%. The friction coefficient of composites containing particles of intermediate hardness is in range II, which is intermediate between ranges I and III

- Formation of a solid lubricating film during sliding wear has been considered as the primary factor responsible for low wear and friction in composites containing soft particles of solid lubricating materials
- Under abrasive wear, the wear rate of aluminum-base composites containing hard particles usually decreases with an increase in volume fraction of particles
- The erosive wear rate in iron- and cobalt-base composites increases with carbide particle content when tested with quartz erodent
- The wear rate is inversely proportional to the particle size in composites containing either hard or soft particles, under either adhesive or abrasive wear conditions
- The wear rate increases with normal load in hard-particle composites under adhesive or abrasive wear. For composites containing higher amounts of soft particles, the wear rate tends to decrease with increasing load up to a critical value before severe wear sets in
- Variations in contact conditions affect wear results. In addition, the relative orientation of the fiber and the sliding direction are important parameters influencing wear
- For metal-matrix composites containing hard particles, sliding against a steel counterface results in generation of steel filings due to microcutting; these filings are compacted during sliding to form a transfer layer
- Under abrasive wear, delamination and micromachining mechanisms are responsible for wear in composites containing hard particles under both two-body and three-body wear conditions

References

1. J.K. Lancaster, Composites for Increased Wear Resistance: Current Achievements and Future Prospects, *New Directions in Lubrication, Materials, Wear and Surface Interactions--Tribology in the 80's,* W.R. Loomis, Ed., Noyes, 1983, p 320
2. P.K. Rohatgi, Y. Liu, and R. Asthana, Some Issues in the Construction of Wear Mechanism Maps for Metal Matrix Composites, *Tribology of Composite Materials,* P.K. Rohatgi, P. Blau, and C. Yust, Ed., ASM International, 1990, p 69
3. I.J. McColm, *Ceramic Hardness,* Plenum Press, 1990
4. *CRC Handbook of Tables for Applied Engineering Science,* 2nd ed., R.E. Bolz and G.L. Tuve, Ed., CRC Press, 1981
5. S.V. Prasad, C.S. Narendranath, and P.K. Rohatgi, Tribological Properties of Al Alloy Matrix Particulate Composites, *Aluminum Alloys--Physical and Mechanical Properties,* E.A. Starke, Jr. and T.M. Sanders, Jr., Ed., EMAS Ltd., 1986, p 1067
6. P.K. Rohatgi, R. Asthana, and S. Das, *Int. Met. Rev.,* Vol 31, 1986, p 115
7. S. Ray, *Ind. J. Technol.,* Vol 28, 1990, p 368
8. P.R. Prasad, S. Ray, J.L. Gaindhar, and M.L. Kapoor, *Scr. Metall.,* Vol 19, 1985, p 1019
9. V.C. Nardone and K.M. Prewo, *Scr. Metall.,* Vol 20, 1986, p 43
10. S.V. Nair, J.K. Tien, and R.C. Bates, *Int. Met. Rev.,* Vol 30, 1985, p 275
11. W.A. Logsdon and P.K. Liaw, *Eng. Fract. Mech.,* Vol 24, 1991, p 757
12. Y.M. Pan, M.E. Fine, and H.S. Cheng, Wear Mechanisms of Al-based Metal Matrix Composites Under Rolling and Sliding Contacts, *Tribology of Composite Materials,* P.K. Rohatgi, P. Blau, and C. Yust, Ed., ASM International, 1990, p 93
13. M.D. Skibo, Report SAND81-8212, Sandia National Laboratories, 1981

14. A. Banerji and P.K. Rohatgi, *J. Mater. Sci.,* Vol 17, 1982, p 335
15. B.C. Pai and P.K. Rohatgi, *J. Mater. Sci.,* Vol 13, 1978, p 329
16. J. Yang and D.D.L. Chung, *Wear,* Vol 135, 1989, p 53
17. D. Nath and P.K. Rohatgi, *J. Mater. Sci.,* Vol 16, 1981, p 1599
18. *Metals Handbook,* 9th ed., Vol 2, *Properties and Selection: Nonferrous Alloys and Pure Metals,* American Society for Metals, 1979
19. H. Hahn, A.D. Divecha, P.J. Lare, and R.A. Hermann, *Material Technology--An Interamerican Approach,* American Society of Mechanical Engineers, 1968, p 231
20. F.M. Hosking, F.F. Portillo, R. Wunderlin, and R. Mehrabian, *J. Mater. Sci.,* Vol 17, 1982, p 477
21. M. Suwa, K. Komuro, and K. Soeno, *J. Jpn. Inst. Met.,* Vol 42, 1978, p 1034
22. D.M. Schuster, M.D. Skibo, and W.R. Hoover, *Light Met. Age.,* Vol 2, 1989, p 15
23. A.D. Divecha and S.G. Fishman, *SAMPE Q.,* Vol 4, 1981, p 40
24. T.P. Murali, S.V. Prasad, M.K. Surappa, and P.K. Rohatgi, *Wear,* Vol 80, 1982, p 149
25. M.K. Surappa, S.V. Prasad, and P.K. Rohatgi, *Wear,* Vol 77, 1982, p 295
26. P.K. Ghosh and S. Ray, *J. Mater. Sci.,* Vol 21, 1986, p 1667
27. R.J. Sample, R.B. Bhagat, and M.F. Amateau, *Proceedings on Cast Reinforced Metal Composites,* ASM International, 1988, p 179
28. Y. Flom and R.I. Arsenault, *Acta Metall.,* Vol 37, 1989, p 2413
29. A. Mortensen, "A Review of the Fracture Toughness of Particle Reinforced Al Alloys," paper presented at Conference on Fabrication of Particulates Reinforced Metal Composites (Montreal), ASM International, Sept 1990
30. S. Ray, *Proceedings on Cast Reinforced Metal Composites,* S.G. Fishman and A.K. Dhingra, Ed., ASM International, 1988, p 77
31. P.K. Ghosh and S. Ray, *J. Mater. Sci.,* Vol 22, 1987, p 4077
32. N. Saka, Effect of Microstructure on Friction and Wear of Metals, *Fundamentals of Tribology,* N.P. Suh and N. Saka, Ed., MIT Press, 1978, p 135
33. A. Sato and R. Mehrabian, *Metall. Trans.,* Vol 713B, 1976, p 443
34. M. Suwa, K. Komuro, and K. Soeno, *J. Jpn. Inst. Met.,* Vol 40, 1976, p 1074
35. B.N. Keshavaram, P.K. Rohatgi, R. Asthana, and K.G. Sathyanaragana, Friction and Wear Behavior of Al Alloy-Glass Particulate Composites, *Tribology of Composite Materials,* P.K. Rohatgi, P. Blau, and C. Yust, Ed., ASM International, 1990, p 133
36. F. Rana and D.M. Stefanescu, *Metall. Trans.,* Vol 20A, 1989, p 1564
37. D. Nath, P.K. Rohatgi, and T.K.G. Namboodhiri, *Tribology of Composite Materials,* P.K. Rohatgi, P. Blau, and C. Yust, Ed., ASM International, 1990
38. D. Nath, S.K. Biswas, and P.K. Rohatgi, *Wear,* Vol 60, 1980, p 61
39. J.D. Ayers, T.R. Tucker, and R.C. Bowers, *Scr. Metall.,* Vol 14, 1980, p 549
40. P.H. Bowen, D.J. Boes, and J.R. McDowell, *Machine Des.,* Vol 8, 1963, p 139
41. Y. Tsuya, *Jpn. Inst. Compos.,* Vol 11, 1985, p 127
42. P.K. Rohatgi, Y. Liu, M. Yin, and T.L. Barr, *Mater. Sci. Eng.,* Vol A123, 1990, p 213
43. P.R. Gibson, A.J. Clegg, and A.A. Das, *Wear,* Vol 95, 1984, p 193
44. L.B. Johnson and D. Kuhlmann-Wilsdorf, *Mater. Sci. Eng.,* Vol 58, 1983, p L1
45. P.K. Rohatgi, S. Ray, and Y. Liu, Friction and Wear Behavior of Metal Matrix-Graphite Particle Composites, *Int. Mater. Rev.,* 1990, accepted for publication
46. E. Yuasa, T. Morooka, and F. Hayama, *J. Jpn. Inst. Met.,* Vol 50, 1986, p 1032
47. J.L. Johnson and J. Schreurs, *Wear,* Vol 78, 1982, p 219
48. L.B. Johnson, Jr. and D. Kuhlmann-Wilsdorf, *Mater. Sci. Eng.,* Vol 58, 1983, p 4
49. Y. Tsuya, *J. Jpn. Inst. Compos.,* Vol 11, 1985, p 127

50. K.C. Owen, M.J. Wang, C. Persad, and Z. Eliezer, *Wear,* Vol 120, 1987, p 117
51. S. Bhattacharrya and F.C. Bock, *Wear,* Vol 46, 1978, p I
52. N.P. Suh, *Tribophysics,* Prentice-Hall, 1986, p 264
53. A.T. Alpas and J.D. Embury, *ASME Conference on Wear of Materials,* American Society of Mechanical Engineers, 1991, p 159
54. K.J. Bhansali and R. Mehrabian, *J. Met.,* Vol 34, 1982, p 30
55. T. Tawarada, S. Okaniwa, and K. Nagahashi, U.S. patent No. 3877884, 1975
56. S.V. Prasad, P.K. Rohatgi, and T.H. Kosel, *Mater. Sci. Eng.,* Vol 80, 1986, p 213
57. N. Saka and D.P. Karalekas, Friction and Wear of Particle Reinforced Metal-Ceramic Composites, *Wear of Materials,* 1985, p 784
58. N.P. Suh, *Wear,* Vol 25, 1973, p 111
59. S. Jahanmir and N.P. Suh, *Wear,* Vol 44, 1977, p 17
60. S. Jahanmir, E.P. Abrahamson, and N.P. Suh, *Wear,* Vol 40, 1976, p 75
61. W.G. Halley, R.E. Swanson, and N.S. Eiss, Abrasive Wear of Particle Reinforced Tool Steels, *Tribology of Composite Materials,* P.K. Rohatgi, P. Blau, and C. Yust, Ed., ASM International, 1990
62. M. Chen, T.Z. Kattamis, B.V. Chambers, and J.A. Cornie, Processing and Abrasive Wear Resistance of Cast TiC-Ferrous Matrix Particulate Composites, *Proceedings of Conference on Engineered Materials for Advanced Friction and Wear Applications* (Gaithersburg, MD), ASM International, 1988, p 1
63. H.C. Sin, N. Saka, and N.P. Suh, *Wear,* Vol 95, 1984, p 193
64. Y. Liu, R. Asthana, and P.K. Rohatgi, *J. Mater. Sci.,* Vol 26, 1991, p 93
65. A.J. Ninham and A.V. Levy, *Wear,* Vol 121, 1988, p 347
66. S.S. Aptekar and T.H. Kosel, *Wear of Materials,* K.C. Ludema, Ed., American Society of Mechanical Engineers, 1985, p 677
67. A. Ball and A.W. Patterson, *Proceedings of the 11th International Plansee Seminar,* Sinter Metall Werke (Austria), Vol 2, 1985, p 377
68. I.G. Wright, D.K. Shetty, and A.H. Clauer, *Erosion by Solid and Liquid Impact,* J.E. Field and N.S. Corney, Ed., Cambridge University Press, Sept 1983, p 63
69. K.C. Goretta, W. Wu, J.L. Routbort, and P.K. Rohatgi, Solid-Particle Erosion of Al/Particulate Ceramic Composites, *Tribology of Composite Materials,* P.K. Rohatgi, P.J. Blau, and C.S. Yust, Ed., ASM International, 1990, p 147
70. M. Suwa, "Hitachi Graphite-Dispersed Cast Alloy-Gradia," Hitachi Report, 1986
71. V.M. Desai, C.M. Rao, T.H. Kosel, and N.F. Fiore, *Wear,* Vol 94, 1984, p 89
72. S. Mohan, V. Agarwala, and S. Ray, *Wear,* Vol 140, 1990, p 83
73. S. Suzuki, H. Hino, Y. Miyashita, and T. Miyata, UK patent No. 2112813A, 1983
74. W.K. Choo and C.H. Hoag, *J. Korean Inst. Met.,* Vol 17, 1979, p 474
75. A. Banerji, S.V. Prasad, M.K. Surappa, and P.K. Rohatgi, *Wear,* Vol 82, 1982, p 141
76. K. Kuniya, H. Arakawa, and T. Namekawa, *Trans. Jpn. Inst. Met.,* Vol 28, 1987, p 238
77. M.R. Jolly, *Foundryman,* Vol 11, 1990, p 509
78. L. Cao, Y. Wang, and C.K. Yao, *Wear,* Vol 140, 1990, p 273
79. M.F. Amateau, R.H. Flowers, and Z. Eliezer, *Wear,* Vol 54, 1979, p 175
80. Z. Eliezer, V.D. Khanna, and M.F. Amateau, *Wear,* Vol 51, 1978, p 169
81. K. Akutagawa, H. Ohtsuki, J. Hasegawa, and M. Miyazaki, Technical Paper No. 870441, Society of Automotive Engineers, 1987
82. A.P. Monteiro Baptista, A.B. de Magahaes, and R.N.J. Duarte, *Proceedings on Cast Reinforced Metal Composites,* ASM International, 1988, p 361
83. M.F. Amateau, W.W. French, and D.M. Goddard, Report No. ATR-75(9450)-3, Aerospace Corporation, 1975

84. S.K. Verma and J.L. Dorcic, *Proceedings on Cast Reinforced Metal Composites,* ASM International, 1988, p 115
85. R. Arikan and S. Murphy, *Wear,* Vol 143, 1991, p 149
86. C.P. You, N.T. Donlan, and J.M. Boileau, Microstructural Observations Associated with the Sliding Wear of an Al Based Composite, *Tribology of Composite Materials,* P.K. Rohatgi, P.J. Blau, and C.S. Yust, Ed., ASM International, 1990, p 157

Friction and Wear of Ceramics

W.A. Glaeser, Battelle Memorial Institute

Introduction

CERAMICS, which are inert nonmetallic solids, and cermets, which are metallic bonded ceramics, have been used for tribological purposes for decades. Although high manufacturing costs have limited their applicability, they have been used in special applications, such as high-temperature or highly corrosive environments, and in situations that require high resistance to wear. The toughened ceramics and fiber-reinforced composites that have recently been introduced have opened up new areas of applicability.

The general class of ceramics (that is, pottery, tiles, bricks, toilets, insulators, and porcelains) can be divided into two product lines:

- Fired clays consisting primarily of small crystals of hydrated aluminosilicates
- Expensive high-tech pure dense oxides, nitrides, carbides, and borides

These high-tech products are used in tribological applications. Whereas the fired clays consist of heterogeneous mixtures of crystals embedded in a glassy matrix, the high-tech ceramics are derived from manufacturing abrasives. These crystalline abrasive materials are refined from the impure grit into high-purity products and are compacted into solids which approach the theoretical density of the material and have very little, if any, glassy binder. The porosity of high-tech ceramics is less than 0.5%, much less than the 5 to 15% porosity present in traditional ceramics. The improvement in structural properties of ceramics continues and is typified by the development of hot isostatically pressed (HIP) silicon nitride balls for applications such as high-precision machine tool bearings.

Applications

Ceramics are used for high-temperature bearings and in very aggressive environments. They are also being substituted for conventional metal alloys in some applications. Examples of current application areas are described below.

Bearings. Sapphire is used for jewel bearings in watches and instruments, whereas silicon nitride is used in some ball bearings, as well as in high-temperature heat engines.

Mechanical Seals. Alumina and silicon carbide are used as seal rings in face seals and floating ring seals, particularly in corrosive and erosive fluids. They are applied in oil fields, the chemical processing industry, and pumps.

High-speed machine tools utilize ceramics because of their resistance to heat, which enables deeper cuts and the use of higher cutting speeds than conventional high-speed steel tools. Silicon nitride, Sialon, Alumina, whisker-reinforced alumina, and titanium carbide/nitride are examples of ceramics that are used for cutting tools. Ceramic tools can successfully and rapidly cut cast iron, which can be very abrasive to ordinary tools.

Guides and Rollers. Alumina is used in the textile industry as the fiber guide for high-speed thread spinning. It is also used for computer tape guides and capstans. Silicon nitride is used for the rollers in furnace conveyors, such as those used in metal heat treating and porcelain firing.

Wear plates made of alumina are used in the handling of abrasive materials, such as coal, ash, and slurries, which are found in coal slurry pumping systems and in cyclones that separate coal and water prior to burning. Paper-handling machinery also utilizes ceramic components.

Advanced Heat Engines. Low-heat-loss engines that can conserve fuel are being designed with ceramic (silicon nitride) cylinders, valves, and piston rings. The intent is to operate at a combustion temperature that is much higher than that of conventional engines and to approach adiabatic conditions. Ceramic parts are also being introduced into engine design to reduce wear.

Medical prostheses that utilize alumina ceramics include bones, joints, and tooth roots.

Ceramic dies made of alumina, zirconia, or Sialon are used for some metal extrusion and drawing applications to provide smoother product surface finishes and longer die life.

Mechanical and Physical Properties

The properties of select commercial ceramics that are typical of those used in tribological applications are given in Table 1. A tool steel is included for comparison purposes. The values given in Table 1 are primarily intended to indicate averages of property levels, rather than actual values. Mechanical values vary in the literature, because ceramics properties are significantly influenced by manufacturing processes. Thus, one grade of ceramic can vary in hardness and toughness, depending on its manufacturer. Important data, such as thermal shock resistance, flexural strength, creep at elevated temperatures, and oxidation or decomposition data, are available only for certain types of ceramics.

Table 1 Typical mechanical and physical properties of generic ceramics

Property	Al_2O_3	SiC	Si_3N_4	PSZ	Sialon	TiC	WC/Co	Pyrex	Steel (01 grade, heat treated)
Density, g/cm^3	3.9	3.1	3.3	5.8	3.2	4.9	3	2.5	7.86
Thermal conductivity, W/m · K (Btu/ft · h · °F)	35.6 (20.6)	126 (73)	27 (16)	1.8 (1.0)	21 (12)	35 (20)	1.6 (0.9)	1.5 (0.87)	51.9 (30.1)
Coefficient of thermal expansion, μm/m · °C	7.1	4	2.3	10.1	3	8	6	5	12.1
Heat capacity, J/kg · K	880	2800	810	400	620	600	. . .	500	448
Hardness, HV · kgf/mm^2	1500	2700	1300	1600	1780	3000	1500	. . .	600
Young's modulus, GPa (10^6 psi)	370 (55)	410 (60)	276 (40)	200 (30)	290 (42)	430 (62)	552 (80)	70 (10)	200 (30)
Tensile strength, MPa (ksi)	262 (38)	100 (15)	524 (75)	700 (102)	450 (65)	896 (130)	896 (130)	. . .	2000 (290)
Compressive strength, MPa (ksi)	2600 (380)	. . .	. . .	1850 (270)	3500 (510)	. . .	. . .	. . .	. . .
Flexural strength, MPa (ksi)	380 (55)	550 (80)	660 (95)	690 (100)	. . .	250 (36)	250 (36)	69 (10)	. . .
Fracture toughness, MPa$\sqrt{\text{m}}$ (ksi$\sqrt{\text{in.}}$)	. . .	4.6 (4.2)	4.5 (4.1)	9.5 (8.6)	7.7 (7.0)	. . .	. . .	. . .	14 (12.7)
Poisson's ratio	0.26	0.19	0.24	0.23	. . .	0.19	. . .	0.2	0.3
Melting point, °C (°F)	2015 (3660)	2700 (4890)	1900 (3450)	2700 (4890)	. . .	. . .	. . .	. . .	1600 (2910)

As shown in Table 1, the hardness of ceramics is significantly higher than that of hardened tool steel, whereas ceramic density is lower. The ceramics also have greater strength values than does steel. Many of the ceramics listed in the table are also stiffer than steel and therefore should have greater beam strength per square unit cross section. Furthermore, the lower Poison's ratio and higher Young's modulus of ceramics (versus steel) results in higher contact stresses and smaller contact areas for Hertzian contacts formed on ceramic balls or rollers in bearings, for instance.

Another desirable property that can be achieved in ceramics is hot hardness, which is valuable in high-temperature applications. The hardness-temperature characteristics of ceramics and metallic tool materials are shown in Fig. 1, which indicates the advantage of using ceramics and sintered carbides, versus the use of high-speed steels. Ceramic tools are being used increasingly in manufacturing environments. The coating of sintered carbide tools with TiN (Ref 1, 2) has been an important improvement in tool endurance.

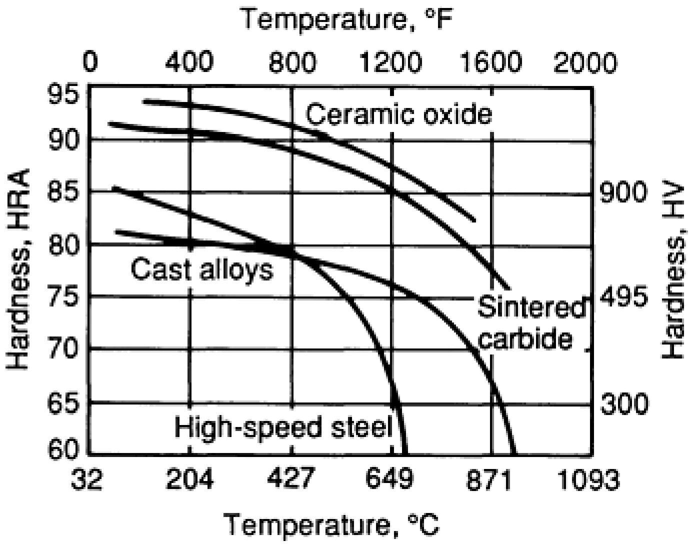


Fig. 1 Hot hardness of tool materials and steels

A large percentage of commercial products are not pure oxide or carbide ceramics. Rather, they are cermets, which are powders bonded by metals or glassy materials. Tungsten carbide and titanium carbide, for example, are generally bonded with cobalt or nickel. The percentage of bonding material in various grades of these cermets will influence their hardness, toughness, thermal conductivity, and abrasion resistance. The effect of cobalt content on tungsten carbide is shown in Table 2. Principally used in tools and rock drilling bits, the cermet metal binder content is adjusted for the potential application. Therefore, it is best to obtain data on properties for a given grade from the manufacturer.

Table 2 Effect of cobalt content on tungsten carbide properties

Cobalt, wt%	Hardness, Rockwell A	Transverse rupture		Young's modulus		Coefficient of thermal expansion	
		MPa	**ksi**	**GPa**	**10^6 psi**	**μm/m · K**	**μin./in. · °F**
3	92.7	1207	175	0.72	0.10	3.0	5.5
6	92	1517	220	0.72	0.10	3.1	5.7
9	89.3	1572	228	0.61	0.09	3.9	7
12	86.8	2069	300	0.58	0.08	4.1	7.5
16	86.8	2503	363	0.52	0.075	4.4	8
25	84.5	2586	375	0.48	0.07	6.1	11.1

As noted previously, ceramics are ideal materials for high-temperature applications that can justify their cost. For example, the hot hardnessvalue of silicon nitride is retained to about 1200 °C (2190 °F). However, some ceramics (other than the oxides) can degrade at elevated temperatures. The carbides and nitrides are subject to oxidation. Those that contain silicon, such as silicon carbide and silicon nitride, depend on a film of silicon dioxide to protect their surfaces from oxidation. Above 1500 °C (2730 °F), the film softens, oxygen diffuses into the bulk material, and unacceptable oxidation of the body occurs. At elevated temperatures, alumina and zirconia, which are fully oxidized, will not oxidize, although the glassy binder in alumina tends to soften, and partially stabilized zirconia undergoes phase changes.

Perhaps the greatest drawback to using ceramics in machinery is their brittleness, or low toughness. Even when advanced toughening techniques are applied, they still fall short of requirements. Ceramics have not yet found acceptance in gas

turbine engines (Ref 3), although operating temperatures have continually increased in new designs. Instead, superalloys are still used with elaborate cooling systems.

The major obstacle to the use of ceramics in advanced gas turbines is an unacceptable fracture toughness. Levels for candidate ceramics and superalloys are shown in Fig. 2. Note that even the toughened ceramics fall below the required maximum fracture toughness level. The advanced low-heat-loss diesel engine, once considered a good candidate for ceramics, has not yet gone into production. High friction losses and thermal shock problems need to be overcome.

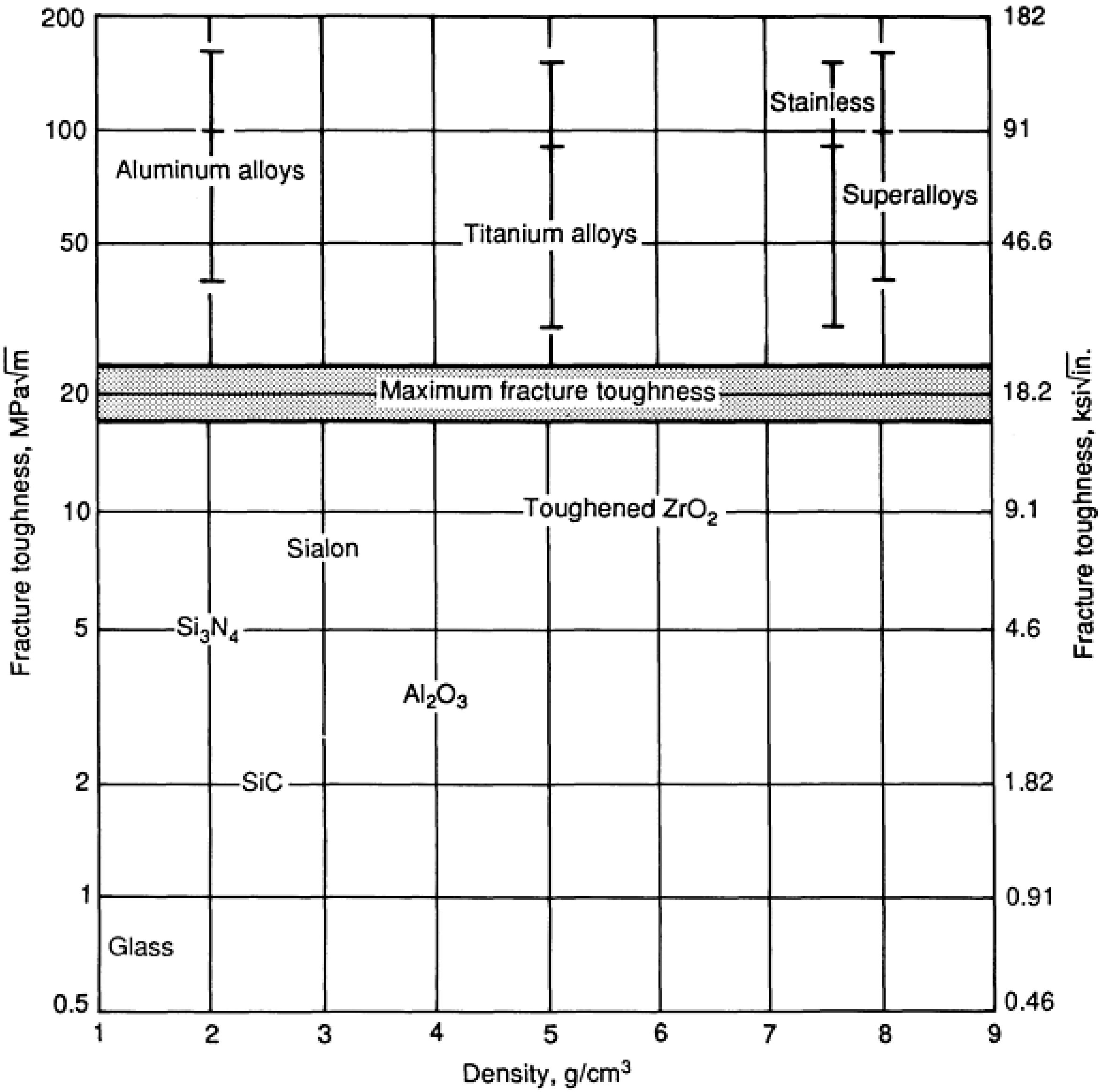


Fig. 2 Comparison of fracture toughness of selected ceramics. Source: Ref 3

Ceramics are being used to a limited extent in heavy-duty diesels, as fuel injector system components. Cam follower rollers represent another development. Both applications utilize silicon nitride. Water pump seals made of silicon carbide are also widely used.

Wear Properties

Because of their inherent brittleness, ceramics can wear by chipping. Surface and subsurface cracks form, join, and release small chips of material. A fine powder is produced as this wear debris is ground up in the wear process. Therefore, ceramics are sensitive to high contact stresses or to any contact condition leading to a state of stress that contains tensile

components. Metals and plastics can deform plastically to relieve high contact stresses before fracture occurs. Ceramics can deform plastically under the hydrostatic stress associated with concentrated contacts, but the plastic deformation involved is very small, when compared with metals and polymers.

Increasing the temperature of a ceramic material to about 0.6 times the melting point, T_m, increases the mobility of dislocations and increases the potential for plastic deformation. Strength reduction and increased creep rates also accompany the rise in temperature. However, the increase in plasticity with increase in temperature does not produce the ductility that is common in metals. Either a brittle or semi-brittle behavior persists.

In instances of sliding contact, brittle materials can exhibit deformation-type wear. In cases of either abrasion or erosion, wear behavior is much different. Both abrasive and erosive processes involve tensile components in the contact state of stress. For abrasive wear, the wear rate is proportional to the elastic modulus and inversely proportional to hardness and fracture toughness. However, it has been shown that alumina and silicon carbides also exhibit a sensitivity to microstructure during wear by scratch tests (Ref 4). The grain size and properties of the intergranular material were found to influence wear.

The thermal-shock-induced fracture of ceramics is a serious wear problem. Because of low thermal conductivity, large thermal gradients can develop in ceramics during frictional heating. When quenched, these hot spots develop large tensile stresses and cracks develop. The result is an increase in wear as relatively large pieces break out of the surface. Eventual failure by fracture can occur.

Ceramics are sensitive to strain rate and will exhibit an increased tendency to crack with increased sliding velocity, in addition to the frictional heating. This sensitivity to impact makes ceramics vulnerable to erosive wear. The angle of impingement of solid-particle erosion influences the rate of material removal. For metals, maximum erosive wear occurs at an impingement angle between 20° and 40°. Ceramics, on the other hand, show high erosive wear when the erosive particle angle nears 90°. Figure 3 shows the difference in erosion rate versus impingement angle for ceramics and metals.

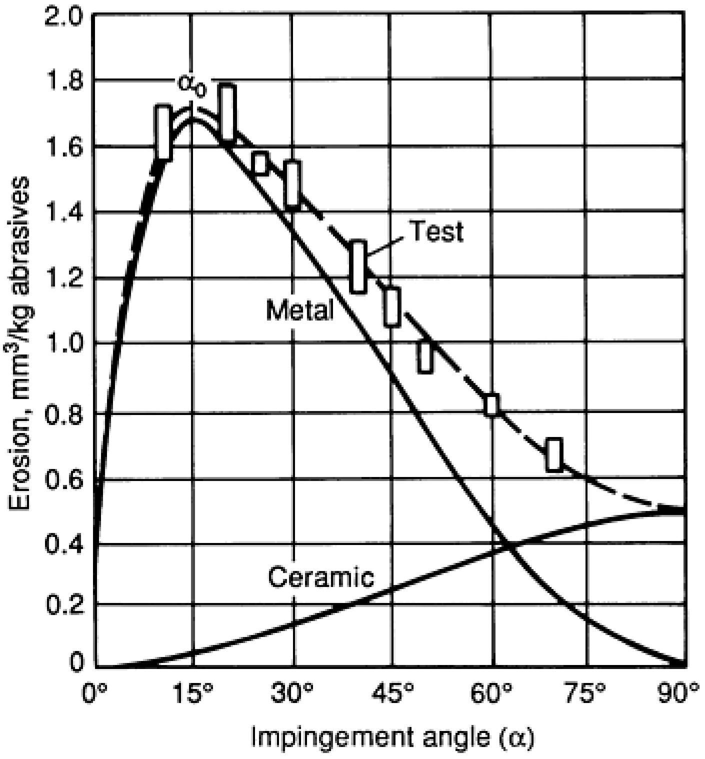


Fig. 3 Effect of erosion impingement angle on material removal rate for ceramics and metals.

The relative hardness of the ceramic and eroding medium is important to erosion mechanisms. When the erodant is either softer or about as hard a the ceramic, a small increase in its hardness causes a large increase in erosive wear. When the

erodant is much harder than the ceramic, microstructure and fracture toughness become more important to the rate of erosion (Ref 5). Thus, a hard ceramic will resist particle erosion as long as the erodant is much softer than it is. if the erodant is harder than the ceramic, then the erosion rate can be minimized by providing as much toughness as possible, a small grain size, and minimum porosity.

Heat-Treatable Ceramics

Transformation-toughened ceramics have been a relatively recent development in which the fracture toughness of certain ceramic "alloys" has been enhanced by taking advantage of phase transformations. Zirconia, in particular, can exist in three different crystal structures: cubic, tetragonal, and monoclinic. The important aspect of this polymorphism is the change in density that accompanies phase transformation. Although zirconia-toughened ceramics are numerous, Y_2O_3 and MgO partially stabilized zirconia (PSZ) are the most developed and most often used materials. Because zirconia has a very high melting point (2850 °C, or 5160 °F), it is potentially a high-temperature structural material. When alloyed with MgO, the phase diagram shown in Fig. 4 (Ref 6) indicates that the structure can be manipulated by heat treatment in a manner similar to that used for metal alloys.

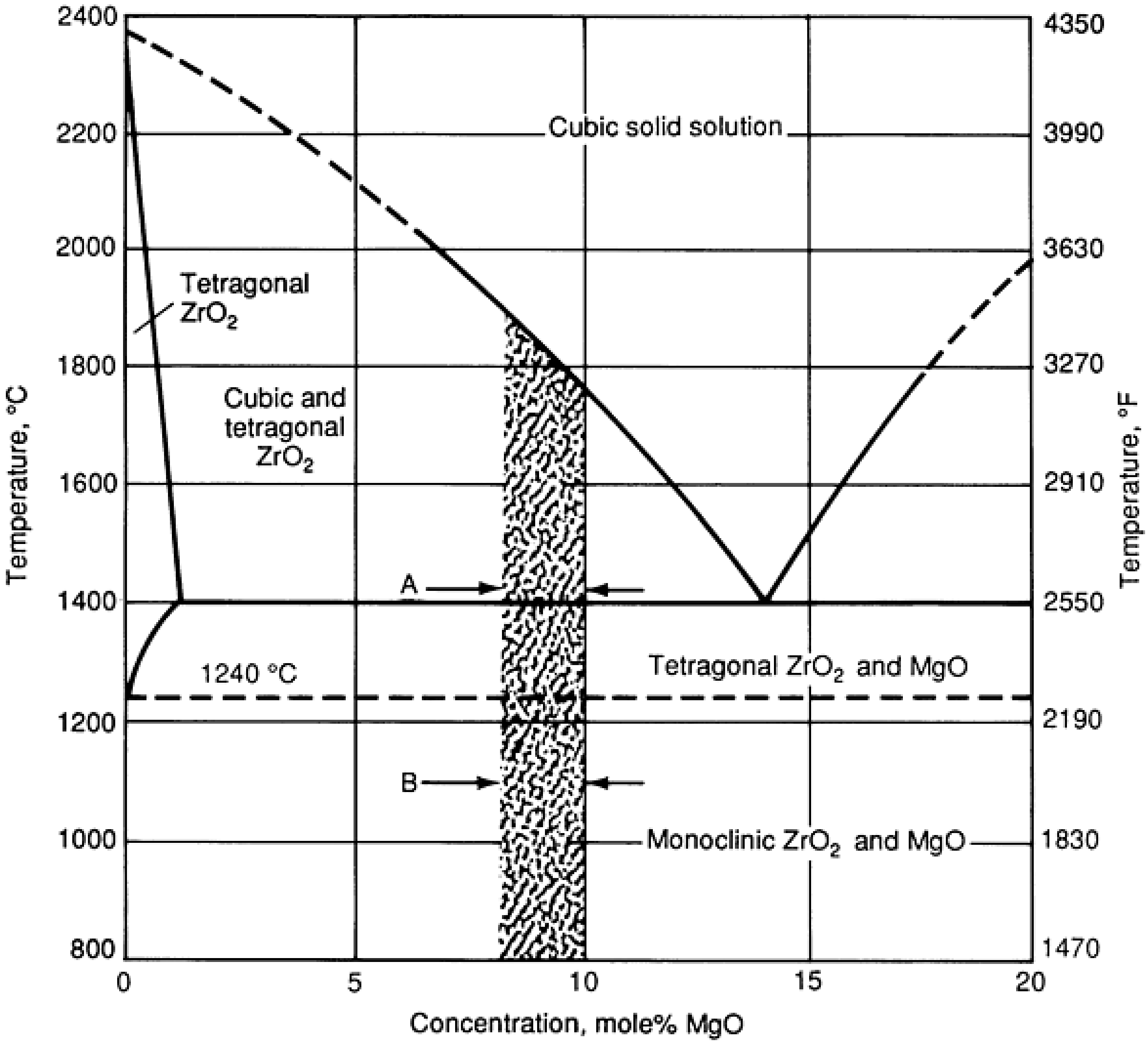


Fig. 4 Zirconia MgO phase diagram. The shaded region indicates the composition range used in most commercially available alloys. Aging is carried out at temperatures above the eutectoid temperature at A to obtain maximum strength and below the eutectoid temperature at B to obtain maximum thermal shock resistance.

The most common MgO-zirconia compositions fall in the shaded range of Fig. 4. The "stabilization" of zirconia, through the addition of MgO, Y_2O_3, or CeO_2, lowers the cubic-to-tetragonal transformation temperature. Without the added stabilizers, this transformation results in tensile stresses and microcracking. Thus, stabilization allows the sintering of zirconia powders in the cubic phase and cooling to room temperature without any phase transformation. Because the transformation shown in Fig. 4 is very sluggish, precipitation hardening is possible. Heating the sintered body to 1800 °C (3270 °F) and then quenching, followed by reheating and annealing at 1400 °C (2550 °F), results in the precipitation of tetragonal particles. If the tetragonal phase is coarsened such that it becomes metastable, then it will transform to the monoclinic phase when stressed by a nucleating crack, and will inhibit crack propagation by producing a compressive field at the crack tip. The heat-treatable zirconia is known as partially stabilized zirconia, which retains hardness and wear resistance while acting as a tougher material.

It has been suggested that frictional heating during sliding contact can also produce phase changes in PSZ (Ref 7). Frictional heating can achieve temperatures that are high enough to transform the monoclinic phase to tetragonal, or even to cubic, resulting in a volume reduction and tensile stress cracking.

Aluminum-oxide ceramics and mullites (alumina-silica) can be toughened by the addition of zirconia (Ref 8), and have been used as advanced tool materials. Other such materials include silicon carbide whisker-reinforced alumina and modified Sialons (Ref 9).

Although zirconia can be toughened and thus successfully used in cam and tappet assemblies, cutting tools, and dies, it is susceptible to thermal shock. Its thermal expansion coefficient is high, and its thermal conductivity is low, which makes it more sensitive to thermal shock than silicon nitride, for example. Thermal diffusivity, or the ability to dissipate temperature gradients rapidly, is low in this material and is affected by thermoelastic instability, as are other low-diffusivity ceramics, such as magnesia and alumina. During sliding contact tests that simulated piston ring-cylinder sliding contact under diesel operating conditions, it was shown that PSZ exhibited classic thermoelastic instability (Ref 10). Self-mated yttria-PSZ that ran dry operated at a high friction coefficient and had a high level of wear, because of thermal shock induced by localized frictional heating.

A critical sliding velocity exists for each material, above which thermoelastic instability (TEI) initiates. An equation has been developed to predict the critical velocity, V_{cr}, for various materials, using their thermal properties (Ref 11). A simplified expression is:

$$V_{cr} = \frac{4K^2}{(\mu \alpha E)^2 \pi k z} \quad \textbf{(Eq 1)}$$

where V_{cr} is the critical sliding velocity, K is thermal conductivity, μ is the coefficient of friction, α is the coefficient of thermal expansion, E is Young's modulus, k is thermal diffusivity ($K/\delta c$ where δ is density and c is specific heat), and z is the width of the slider.

The velocity used in the simulated engine tests was 5.4 m/s (18 ft/s). The coefficient of friction for yttria-PSZ was 0.13, and the estimated V_{cr} was 8.9. For silicon nitride, the coefficient of friction was 0.08, and the estimated V_{cr} was 538. The silicon nitride performed without evidence of TEI.

Rolling Contact Bearings

High-speed ball bearings become subject to a higher incidence of rolling contact fatigue damage as speeds increase. This is the result of centrifugal force between the balls and the outer races. Advanced high-speed gas turbines are limited in performance by this effect. By using a less-dense ball material, such as silicon nitride, this effect can be reduced. It has been demonstrated that five-ball rolling contact tests indicated that low-mass silicon nitride balls performed as well as bearing steel balls (Ref 12). In other words, the fatigue life of hot-pressed silicon nitride balls in the five ball apparatus was equal to that of typical bearing steels.

More recently, continuing development of silicon nitride components in ball bearings for applications such as gas turbines and high-speed spindles for machine tools continue to show promise as solutions for the extreme operating demands of these components. Hybrid bearings having silicon nitride balls and tool steel races have been developed for very high-speed applications. The centrifugal loads on the outer races that result from steel balls used in high-speed bearings can be

reduced by using lower density silicon nitride balls. Endurance tests conducted on these hybrid bearings have determined the contact stress limit to be 2000 MPa (290 ksi), identical to the contact stress limit of all-steel bearings (Ref 13).

The rolling contact fatigue life of hot-pressed silicon nitride (HPSN) balls containing yttria and alumina additives have been found to be greater than conventional ball bearings steel balls (Ref 14).

The continuing research into the use of ceramics in rolling contact bearings for both high-speed and high-temperature applications has produced encouraging successes. Note, however, that these bearings are expensive and must be specially designed by bearing specialists to ensure reliable performance.

References

1. T. Sadahiro, S. Yamaya, K. Shibuki, and N. Ujiie, Wear Resistant Coating of Cemented Carbides and High Speed Steels by Chemical Vapor Deposition, *Wear,* Vol 48, 1978, p 291-299
2. F. Heydari, J. Wallbank, and I.R. Pashby, Evaluation of Wear Mechanisms and Tool Lives When Machining 0.3% Mn-Mo Steel with Coated Carbide and Ceramic Tools, *Proceedings of the International Conference on Wear of Materials 1987,* American Society of Mechanical Engineers, p 313-324
3. C.T. Simms, Non-Metallic Materials for Gas Turbines, *Adv. Mater. Process.,* June 1991, p 32-39
4. O.O. Ajayi and K.C. Ludema, The Effect of Microstructure on Wear Modes of Ceramic Materials, *Proceedings of the International Conference on Wear of Materials 1991,* American Society of Mechanical Engineers, p 307-318
5. R.A. Vaughn and A. Ball, The Effect of Hardness and Toughness on the Erosion of Ceramic and Ultrahard Materials, *Proceedings of the International Conference on Wear of Materials 1991,* American Society of Mechanical Engineers, p 71-75
6. R.H.J. Hannick and M.J. Murray, Magnesia-Partially Stabilized Zirconias as Wear Resistant Materials, *Proceedings of the International Conference on Wear of Materials 1983,* American Society of Mechanical Engineers, p 181-186
7. W. Bundschuh and K.-H Zum Gahr, Influence of Porosity on Friction and Sliding Wear of TZP-Zirconia, *Proceedings of the International Conference on Wear of Materials 1991,* American Society of Mechanical Engineers, p 319-326
8. A.H. Heuer, Transformation Toughening in ZrO_2 Containing Ceramics, *J. Am. Ceram. Soc.,* Vol 70 (No. 10), p 689-698
9. R. Komanduri and S.K. Samanta, Ceramics, *Metals Handbook,* 9th ed., Vol 16, ASM International, 1989, p 98-103
10. K.F. Dufrane, Sliding Performance of Ceramics for Advanced Heat Engines, *Ceram. Eng. Sci. Proc.,* Vol 7 (No. 7-8), 1986, p 1052-1059
11. T.A. Dow and R.A. Burton, The Role of Wear in the Initiation of Thermoelastic Instabilities of Rubbing Contact, *J. Lubr. Technol. (Trans. ASME),* Series F 95 (No. 1), 1973, p 71-75
12. R.J. Parker and E.V. Zarestsky, Fatigue Life of High-Speed Ball Bearings with Silicon Nitride Balls, *J. Lubr. Technol. (Trans. ASME),* July 1975, p 350-357
13. F.J. Ebert, "Performance of Silicon Nitride Components in Aerospace Bearing Applications," Paper 90-GT-166, International Gas Turbine and Aeroengine Congress, American Society of Mechanical Engineers, 1990
14. T. Fujiwara, T. Yoshioka, T. Kitahara, and S. Koisumi, Study on the Load Rating Property of Silicon Nitride for Rolling Bearing Material, *Jpn. Soc. Lubr. Eng. J.,* No. 10, 1989, p 81-86 (in English)

Friction and Wear of Carbon-Graphite Materials

Stanley Chinowsky, Pure Carbon Company

Introduction

CARBON-BASED MATERIALS are widely used in friction and wear applications. They are commercially available in hundreds of grades and have a wide range of physical, chemical, and electrical properties. Tribological characteristics can be tailored for numerous applications by the selection of raw materials and manufacturing processes. The combination of carbon-based materials with metallic, polymeric, fiber, inorganic chemical, and other additives can create infinite variety. In the present state of the art, only past experience and testing can determine the suitability of a material for an application.

Carbon, the major chemical constituent of carbon and graphite materials, can vary in terms of its physical structure. It can be amorphous in glassy carbon, turbostratic in carbon black, imperfect hexagonal crystals in coke and coal, near-perfect hexagonal crystals in graphite, or near-perfect tetragonal crystals in diamond. Carbon fibers, in various degrees of chemical purity and crystalline perfection, add to the available choices.

Generally, a material described as carbon suggests an imperfect-to-amorphous crystal structure, processing temperatures up to about 1100 °C (2010 °F), low conductivity, and a relatively hard and abrasive material. Material described as graphite would have a more-perfect crystal structure, and would have been heated to nearly 3000 °C (5430 °F) during manufacture. It would also have higher conductivity and be soft and slippery. Many materials are intermediate in terms of all of the descriptors, and these hybrids are often described as carbon-graphite.

Carbon and graphite materials are manufactured for numerous applications that have little to do with friction and wear. Applications of this type include arc furnace electrodes, battery anodes, electrical discharge machining electrodes, gouging and welding rods, and continuous casting molds. In some of these applications (electrodes and gouging rods), materials are optimized for minimum wear from erosion and oxidation. In other cases (continuous casting), friction plays a role. However, materials used in these applications are not primarily engineered for friction and wear.

Friction and wear applications of carbon and graphite materials can be classified as being either electrical or mechanical, depending on whether electric current is carried by the carbon rubbing component. Materials are tailored for one or the other type of use; it would be very unusual for a material to find application in both areas. Electrical carbons are used for motor brushes, potentiometer contacts, and other applications in which current must be transferred across a sliding interface. Mechanical carbons are used for sliding members that do not carry current, such as bearings, seals, pump vanes, and rotors.

Electrical carbons are frequently low in strength (less than 35 MPa, or 5.1 ksi, in flexural strength), because forces are generally low and are often in compression. Mechanical carbons are subjected to larger forces that can stress the carbon in tension and are therefore usually stronger and stiffer.

Metals are often combined with carbon to make composite materials with increased conductivity and strength. Copper, silver, lead alloys, and antimony are commonly used for both electrical and mechanical applications. The combination can be accomplished by either mixing powders before compaction of the material or infiltrating a porous carbon body with a molten metal.

Organic and inorganic compounds of all kinds can also be combined with carbon to influence various properties. As with metals, such additions can be made during powder mixing or by the infiltration of liquid compounds or solutions. Fluorine compounds and molybdenum disulfide are used to improve friction and wear in low-moisture environments. Phosphorus and boron compounds can improve resistance to oxidation. Oils and waxes can be used to modify friction. Possible additives to carbon are almost limitless, because carbon is inert to most reagents. Usually, additives are combined physically with the carbon base material, with no chemical interaction.

Because carbon and graphite materials are produced by pyrolysis of hydrocarbon binders, they are normally porous. The porosity makes it possible to infiltrate materials, as discussed above, to modify properties, but to also leave the materials permeable to fluids. In applications where permeability must be limited or eliminated, such as seals, the pores are filled with an impermeable material, usually a polymer. Phenolic, polyester, and epoxy resins are all commonly used. Applications, such as air bearings, can also make use of the inherent porosity of carbon and graphite.

The addition of other materials to carbon makes the resultant product more susceptible to environmental deterioration than the base carbon itself. Because almost any additive has a lower temperature capability than the carbon itself, the additive limits the operating temperature. Oxidation is an exception. Similarly, because almost any additive is more reactive chemically than carbon, thermal oxidation excepted, the additive determines resistance to attack by corrosive reagents.

Manufacturing Methods

Conventional carbon manufacturing technology uses particulate fillers with either liquid or thermoplastic binders. A mixing process disperses the fillers in the binder and coats the filler particles in the process. The binder is fluid during the mixing process, either through dissolution (as in some resin systems) or by heating (used with pitch or solvent-free resins). The mix can be extruded directly into long cylinders, or it can be further processed to make powder for molding via several different processes, such as extrusion, uniaxial compression in dies in mechanical or hydraulic presses, or isostatic pressing.

After molding, the green compacts are baked to about 1100 °C (2010 °F) to carbonize the binder. For some materials, the initial baking or carbonizing is followed by higher-temperature heat treatments, such as up to 3000 °C (5430 °F) for some graphites. Baked and heat-treated blanks or billets are machined and treated by various infiltration processes to achieve final composition and properties.

Fiber-reinforced carbon materials (carbon-carbon composites) are made with carbon fiber fillers, rather than particulate fillers. The manufacturing technology for both particulate and fiber-filled carbon is similar to methods used for polymer composites, where processing is complete after the binder is cured. The presence or absence of the carbonization process is what differentiates polymer composites from carbon and graphite materials.

Physical and Mechanical Properties

When compared with metals, particulate carbons are low-density brittle materials. Their values for modulus of elasticity and tensile strength are low, but their compressive strength is higher than tensile strength. Thermal conductivity varies widely with the type of carbon. The limiting temperature for practical use is determined by either oxidation rate or changes in the materials used for impregnation.

Table 1 shows how properties can vary from a pure carbon to a pure graphite composition. Fully graphitic material is much higher in thermal conductivity and oxidation resistance, at the expense of hardness and strength. Table 2 shows how the properties of a carbon-graphite are altered by impregnation with a polymer, a metal, and an inorganic salt. The treatments increase strength and density in particular, As discussed in the section "Friction and Wear" below, they also modify tribological behavior.

Table 1 Properties of typical carbon and graphite materials

Property	Carbon (100%)	Carbon-graphite (70%-30%)	Graphite-carbon (70%-30%)	Graphite (100%)
Apparent density, g/cm^3	1.70	1.72	1.75	1.80
Hardness, scleroscope	100	85	65	40
Compressive strength, MPa (ksi)	300 (43)	208 (30)	145 (21)	55 (8)
Flexural strength, MPa (ksi)	62 (9)	62 (9)	52 (7.5)	28 (4)
Modulus of elasticity, GPa (10^6 psi)	21 (3)	17 (2.4)	14 (2.1)	10 (1.4)
Thermal conductivity, W/m · K (Btu · ft/h · ft^2 · °F)	5 (3)	9 (5)	12 (7)	85 (50)
Temperature limit (air), °C (°F)	315 (600)	315 (600)	315 (600)	455 (850)

Table 2 Impregnation effects on typical carbon-graphite base material

Property	Not impregnated	Phenolic resin impregnated	Silver metal impregnated	Lithium fluoride impregnated
Apparent density, g/cm^3	1.72	1.80	2.40	1.85
Hardness, scleroscope	85	90	95	90
Compressive strength, MPa (ksi)	280 (30)	241 (35)	310 (45)	276 (40)
Flexural strength, MPa (ksi)	62 (9)	76 (11)	93 (13.5)	86 (12.5)
Modulus of elasticity, GPa (10^6 psi)	17 (2.4)	21 (3.1)	27 (3.9)	21 (3)
Thermal conductivity, W/m · K (Btu · ft/h · ft^2 · °F)	9 (5)	9 (5)	14 (8)	9 (5)
Temperature limit (air), °C (°F)	315 (600)	260 (500)	290 (550)	315 (600)

Table 3 shows similar effects for a graphite-base material. Note that this graphite is about twice as strong as the graphite listed in Table 1, illustrating the great dependency of properties on materials and processes used for specific compositions.

Table 3 Impregnation effects on typical graphite-base material

Property	Not impregnated	Phenolic resin impregnated	Silver metal impregnated	Carbon impregnated
Apparent density, g/cm^3	1.82	1.90	2.40	1.88
Hardness, scleroscope	75	80	80	80
Compressive strength, MPa (ksi)	138 (20)	172 (25)	200 (29)	145 (21)
Flexural strength, MPa (ksi)	55 (8)	66 (9.5)	76 (11)	61 (8.8)
Modulus of elasticity, GPa (10^6 psi)	12 (1.8)	14 (2.0)	17 (2.4)	14 (2.1)
Thermal conductivity, W/m · K (Btu · ft/h · ft^2 · °F)	70 (40)	60 (35)	85 (50)	70 (40)
Temperature limit (air), °C (°F)	455 (850)	260 (500)	480 (900)	455 (850)

Fiber-reinforced carbons can be much stronger and stiffer than particulate carbons, but properties may be strongly anisotropic. The properties of particulate carbons are usually relatively isotropic. Coarse-grained extruded carbon can show marked anisotropy, but such materials are not usually applied to tribological areas. The thermal expansion of particulate mechanical carbons is lower than it is for metals and polymers, and is comparable to that of ceramics.

Friction and Wear

The tribological characterization of carbon and graphite materials is commonly done under dry or boundary lubrication conditions. Performance is influenced by environment, mating material, surface finish, load, and speed. Friction and wear must be characterized for a mating pair, that is, for a carbon-mating material combination, rather than the carbon itself.

Figure 1 illustrates the great effect that mating material has on the apparent behavior of a specific carbon in dry wear tests. The mating material effects not only the magnitude of the carbon wear, but also its temperature dependence. Note that wear is expressed in a dimensionless unit obtained by dividing the wear depth by the distance rubbed in the test.

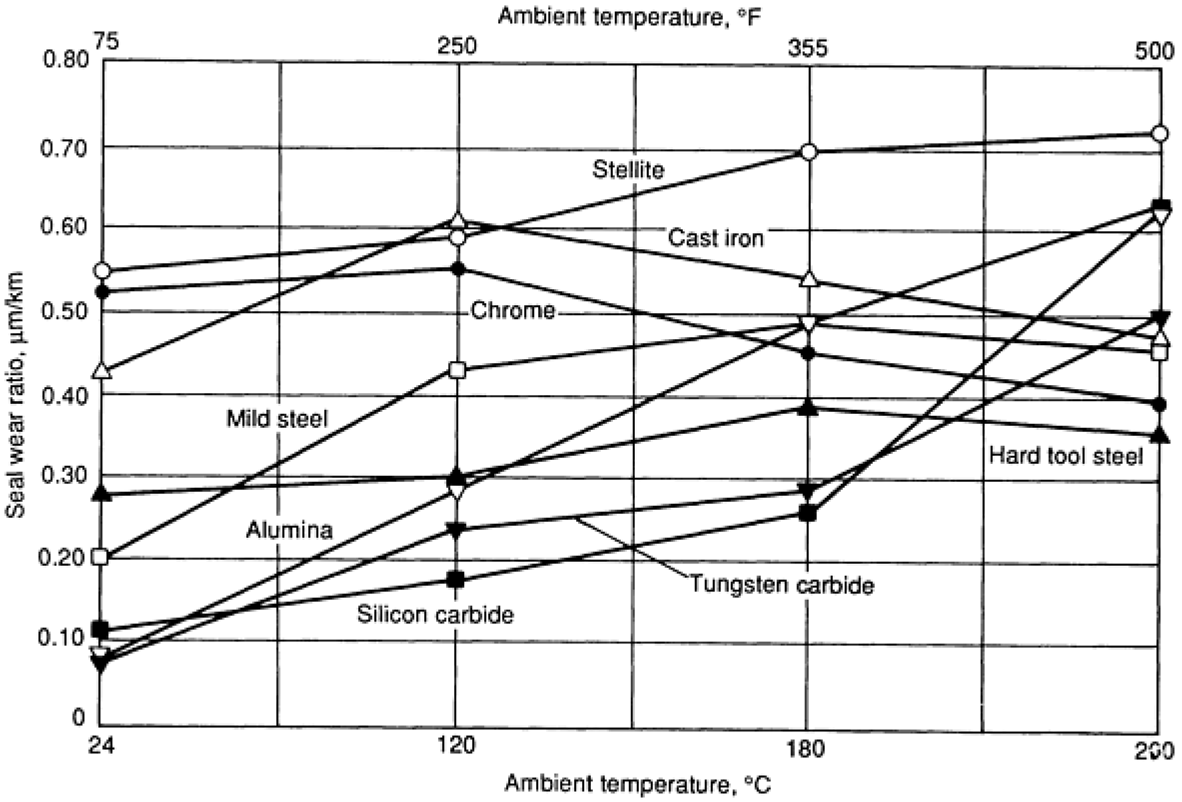


Fig. 1 Wear of a resin-impregnated carbon-graphite against various mating materials. Load, 138 kPa (20 psi); speed, 135 m/min (442 ft/min)

Wear of the mating material is dependent on the carbon rubbing against it. An abrasive, high-ash carbon material can even cause high wear to occur in very hard mating materials, such as tungsten carbide, as shown in Fig. 2.

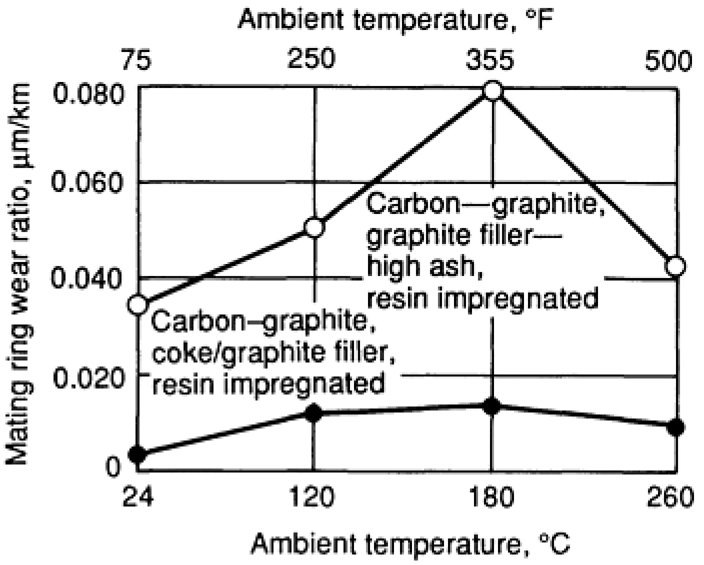


Fig. 2 Mating ring wear with different carbon-graphite materials. Mating material, tungsten carbide; load, 138 kPa (20 psi); speed, 135 m/min (442 ft/min)

Different types of carbon exhibit marked differences in friction and wear for the same conditions. Behavior is often not what might be predicted from physical properties or generic material descriptions. Figure 3 shows that the harder carbon-graphite materials wear more than the relatively soft electrographite material.

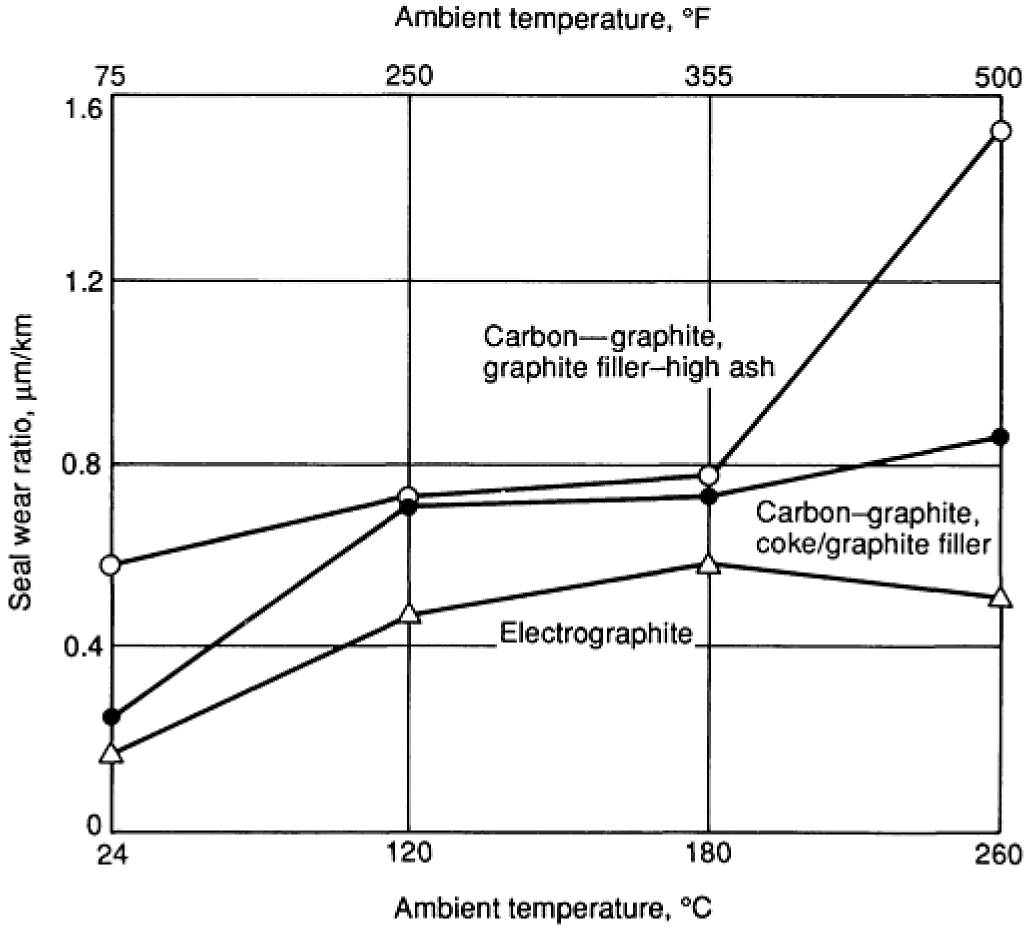


Fig. 3 Wear of different types of carbon. Mating material, SAE 1020 steel; load, 138 kPa (20 psi); speed, 135 m/min (442 ft/min)

Such results apply to the specific grades tested only. It does not follow what all carbon-graphites will wear more than all electrographites under all conditions. Dry rubbing behavior of carbon and graphite materials cannot be predicted in any general or theoretical way. Recommendations can be made for specific cases, based on laboratory tests and field experience. Such recommendations can be made by manufacturers of carbon materials, using their databases from past applications. However, performance should always be verified by the user before final acceptance of the rubbing material system.

Unanticipated results are often found for friction, as well as wear. Generically, one might expect lower friction from a "slippery" electrographite than from a more-abrasive carbongraphite. However, Fig. 4 shows a case where the reverse is true.

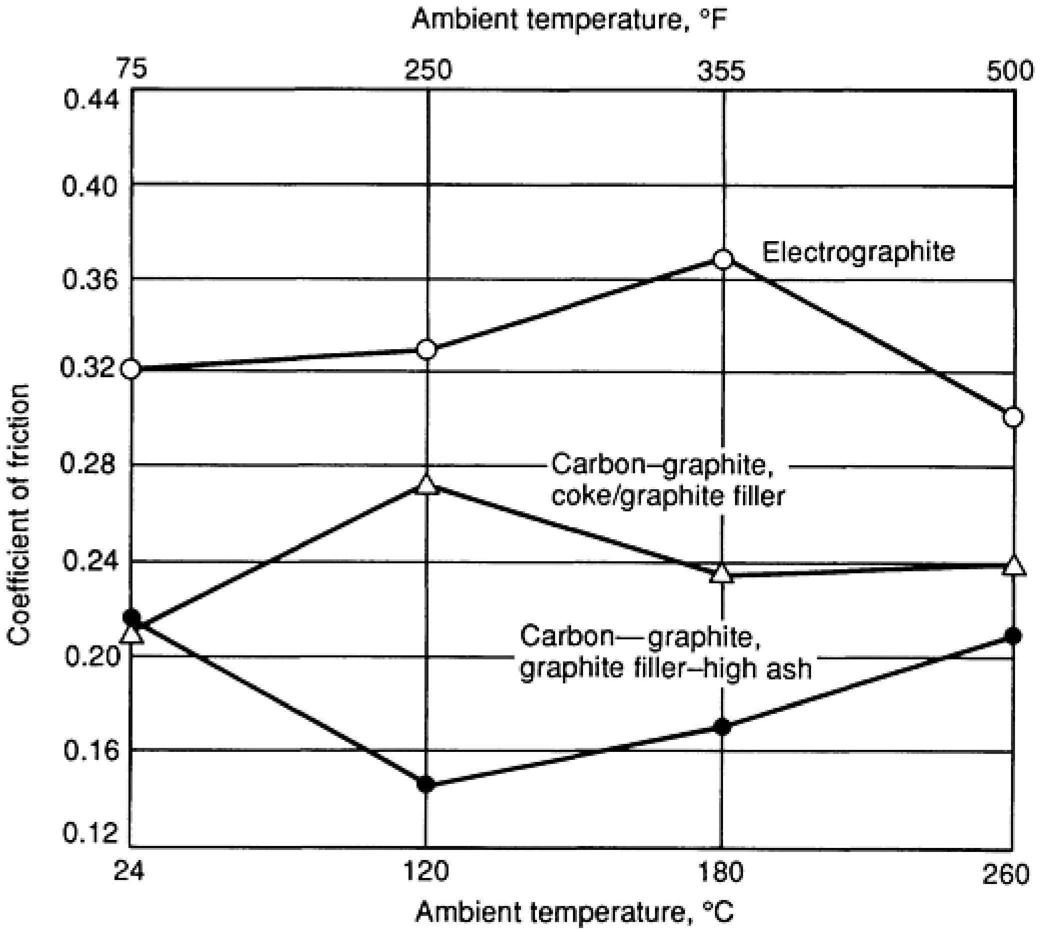


Fig. 4 Friction of different types of carbon. Mating materials, SAE 1020 steel; load, 138 kPa (20 psi); speed, 135 m/min (442 ft/min)

This figure illustrates a segment of carbon and graphite tribology that is sometimes called "film control." The rubbing interface often consists of a thin layer of carbon transferred to the mating ring in the early stages of rubbing. If the transfer film is heavy, as is often the case with graphitic materials, high friction can result. Harden carbons usually deposit a thinner transfer film, generally leading to lower friction.

Filming and other tribological results are influenced by impregnation of basic carbon materials. Figure 5 shows the effects of different types of impregnants on an electrographite.

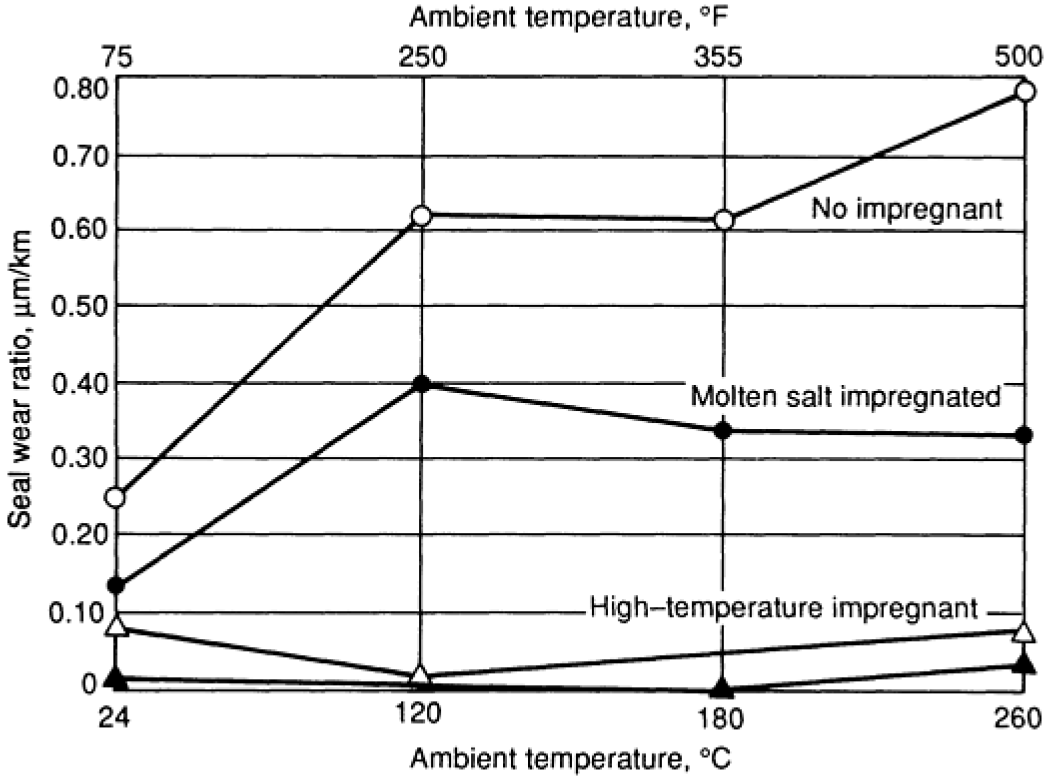


Fig. 5 Wear of an electrographite and related grades. Mating material, chromeplated steel; load, 138 kPa (20 psi); rubbing speed, 135 m/min (442 ft/min)

A similar caveat applies to generalization of the data shown in Fig. 5. The results illustrated were obtained for a specific electrographite with specific impregnants. The same impregnants applied to other electrographites, or to other carbon materials, could produce very different results.

Impregnants influence friction, as well as wear. In one case shown in Fig. 6, a high-temperature impregnant can increase friction at room temperature, but decrease it greatly at higher temperatures.

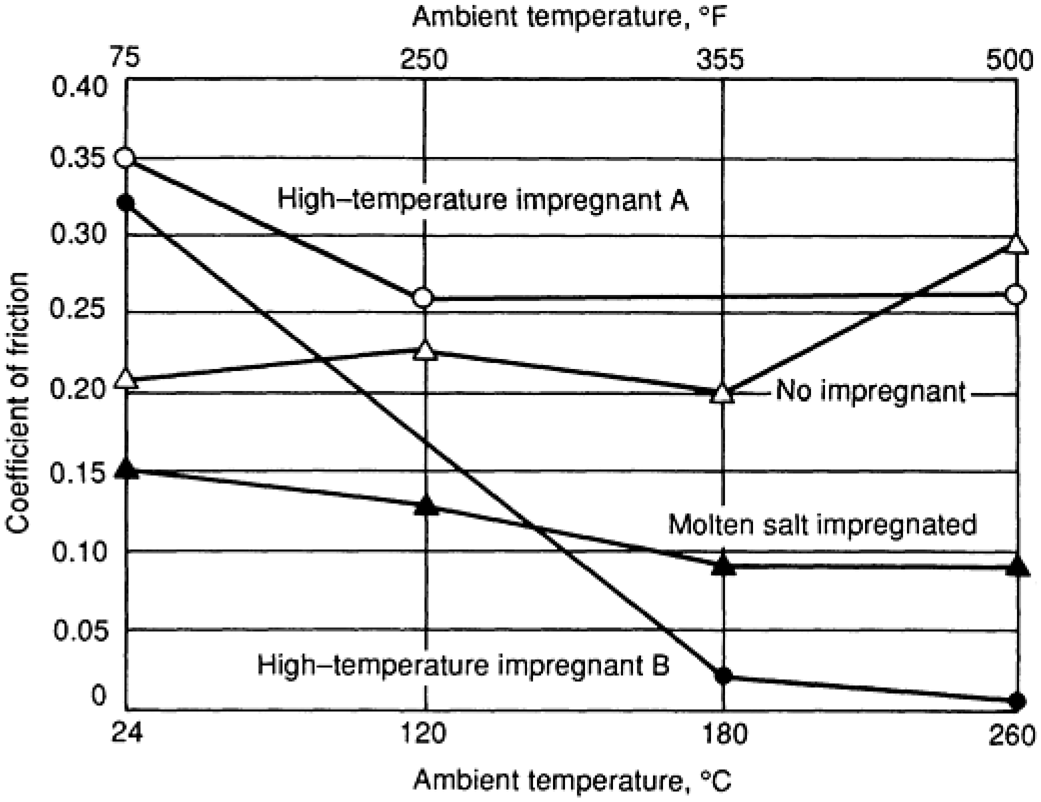


Fig. 6 Friction of electrographite and related grades; same material and conditions as Fig. 5

In boundary and hydrodynamically lubricated applications, wear and friction can be much lower than in the dry rubbing examples presented. Lubrication in these applications in often provided by a film of liquid, such as water, that does not possess lubricating qualities, but does serve to separate parts that are moving relative to each other. The coefficient of friction in seals running with boundary lubrication is typically less than 0.1. In bearings that are hydrodynamically lubricated, the coefficient of friction is less than 0.01, once the hydrodynamic film is formed.

In many applications where the steady state is a lubricated one, the transient start-up and shutdown conditions involve dry rubbing. Carbon and graphite materials are often able to survive the dry running intervals that dictate the selection of carbon for the application. Except for other types of self-lubricating materials, only carbon can be applied without the addition of a lubricating system to protect against failure from dry rubbing.

Although the dry rubbing data illustrates the effects on friction and wear of mating material and temperature, other operating variables are of equal importance. Rubbing speed must be considered, and its effect is not obvious. Some carbon materials will perform better at slow speeds, whereas other materials may run with much lower wear rates at high speeds. For dry rubbing, atmospheric composition is critical, particularly with respect to moisture content. Carbon materials formulated specifically for operation in dry conditions are available. The use of a material that is not suitable for a dry atmosphere may cause the carbon to literally disappear in a cloud of dust.

Under conditions such as cyclic temperature and pressure, some fluids will interact with carbon to cause surface disruptions, called blisters Some grades of carbon resist this failure mechanism better than others. Prior history, laboratory testing, and field verification should all be used in material selection.

Component Design

Performance in specific environments can be strongly influenced by component design. Grooves and indentations of various kinds can be used to balance fluid pressure and reduce effective contact pressure. Grooves can be used in dry applications to aid in wear debris removal, thereby reducing wear.

Mechanical limitations imposed by the strength and modulus of carbon can be dealt with by retaining the carbon in a metal component. Assembly can be accomplished with adhesives, mechanical fasteners, or an interference fit obtained by heating the metal component before assembly. The low modulus of carbon makes the interference fit relatively easy to accomplish, but the resulting distortion of the carbon makes it necessary to finish-machine the carbon after assembly. Because the coefficient of expansion of the metal retainer is always significantly higher than that of the carbon, the temperature capability of the assembly is limited by the amount of interference.

Friction and Wear of Thermoplastic Composites*

Introduction

INJECTION-MOLDABLE THERMOPLASTIC COMPOSITES are increasingly displacing metals in many gear and bearing applications. The ease and economics of manufacturing complex parts by injection molding are well recognized. The ability to absorb shock and vibration, and to operate with less power and noise--with little or no maintenance--are additional advantages gained with thermoplastic composites. Further improvements in mechanical and tribological properties can be attained through the incorporation of glass or carbon fiber reinforcement and polytetrafluoroethylene (PTFE) or other lubricants. These fillers, reinforcements, and lubricants, when compounded with the appropriate base resin, yield an almost limitless array of compounds that provide increasing load-carrying capacity, reduced coefficients of friction, improved wear life, higher mechanical strengths, improved thermal properties, greater fatigue endurance and creep resistance, and excellent dimensional stability and reproducibility.

A large amount of tribological data has been gathered for thermoplastic composites at room temperature. However, continuing tribological characterization of these composites at elevated temperature is required to assist design engineers in expanding their use to more demanding applications. This article focuses on the results of studies performed by the authors on the elevated-temperature friction and wear properties of thermoplastic composites. These results also provide an indication of the overall tribological behavior of polymeric materials.

Previous studies have demonstrated that wear rates and coefficients of friction in plastics versus steel increase with the temperature of the surrounding environment (Ref 1). New opportunities in the transportation industry indicate that increased thermal capabilities for thermoplastic composites are required. As new resins are developed to meet these demands, internally lubricated composites must be developed with corresponding thermal stability and increased resistance to wear at elevated temperatures. To this end, internal lubricant additive with superior thermal stability must be developed and characterized.

The inherent corrosion resistance of plastics is well recognized. This property has become increasingly more important in the chemical process industry and the materials transport industry. Increasing demands are being placed on the structural properties of thermoplastic composites, particularly at elevated temperatures. A large amount of data exist for the structural properties of high-temperature thermoplastic composites. Designers require similar data for the tribological behavior of these composites.

Note

* Adapted from "The Effects of Lubricants and Reinforcements on the Tribological Properties of Thermoplastic Composites at Elevated Temperatures," by Mark P. Wolverton, Kathleen Talley, and John E. Theberge (in *Tribology of Composite Materials,* ASM International, 1990)

Sliding Wear in Thermoplastic Composites

The frictional properties of thermoplastic composites vary in a unique way from metals. When contrasted with metals, even highly reinforced resins have low modulus values and their asperities will more greatly deform under load. The friction between thermoplastic composite and metal surfaces is characterized by adhesion and deformation, resulting in frictional forces that are not proportional to load (that is, the coefficient of friction decreases as load increases) but are proportional to speed (Ref 2).

Wear rate is often defined as the volumetric loss of material over a unit of time. Several mechanisms operate simultaneously to remove material from the wear interface. The primary mechanism is adhesive wear, which is characterized by fine particles of polymer being removed from the surface. The presence of this powder is a good indication that the rubbing surfaces are wearing properly. The presence of melted polymer or large gouges or grooves at the interface is normally an indication that the materials are not wearing by adhesion but rather by abrasion, or an indication that the pressure velocity limit of the material has been exceeded.

The wear of an unlubricated surface in sliding is proportional to the load multiplied by the distance traveled (Ref 3):

$$W \approx FD \qquad \textbf{(Eq 1)}$$

$$W \approx FVT \qquad \textbf{(Eq 2)}$$

The parameters in Eq 1 and 2 are defined in Table 1. Introducing a factor of proportionality (the wear factor), *K*, Eq 2 becomes

$$W = KFVT \qquad \textbf{(Eq 3)}$$

Solving for *K* yields:

$$K = \frac{W}{FVT} \qquad \textbf{(Eq 4)}$$

Table 1 Parameters that affect sliding wear of unlubricated surfaces

Parameter	Definition	Units	
		Metric	**English**
W	Wear volume	mm^3	$in.^3$
F	Force	N	lbf
V	Velocity	m/s	ft/min
T	Elapsed time	s	h

The wear factor for a thermoplastic composite is generated by the thrust washer test procedure (ASTM D 3702). The static and dynamic coefficients of friction are also generated using this test. Once a wear factor, *K*, has been established, it can be used to calculate wear rates of bearings, gears, and so on. However, because wear rate is affected by material type, finish, and hardness, as well as by environmental temperature and part design, large errors may result as end-use variables begin to differ from those selected for the test. As a measure of the performance of one composite relative to another at the same operating conditions, the *K* factors have proven to be highly reliable. The units for the *K* factor in this paper are in the English units most familiar to the design engineers who use these data. A conversion to SI units yields 1×10^{-10} $in.^3 \cdot min/ft \cdot lbf \cdot h = 2.014 \times 10^3$ $mm^3/N \cdot m$.

Wear Tests

Wear Factor. Wear tests are conducted with a thrust washer test apparatus, as described in ASTM D 3702. The sample thrust washer is mounted in an antifriction bearing equipped with a torque arm. The raised portion of the thrust washer bears against a dry, cold rolled, carbon steel wear ring with 0.30 to 0.40 μm (12 to 16 μin.) finish and 18 to 22 HRC

hardness. Each evaluation is conducted with a new wear ring that has been cleaned and weighed on an analytical balance. Weight loss and frictional torque are continuously monitored. Figure 1 is a schematic representation of this test.

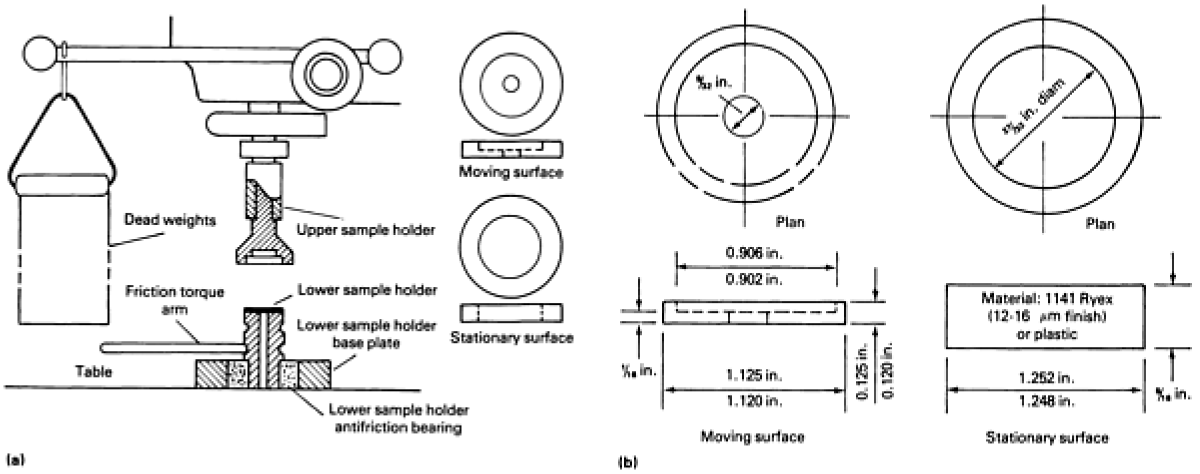


Fig. 1 ASTM D 3702 test used to determine wear rates of self-lubricated materials. (a) Schematic showing thrust washer test apparatus. (b) Schematic showing dimensions of moving test specimen and stationary washer

The test duration depends upon the period required to achieve 360° contact between the raise portion of the thrust washer and the wear ring. The average wear factor and duration of this break-in period are recorded. The wear factors reported for each compound are based on the equilibrium wear rate independent of break-in-wear.

Volume wear is calculated as follows:

$$W = \frac{m}{\rho} \quad \text{(Eq 5)}$$

where W is wear volume in cm^3; m is the weight loss in milligrams; and ρ is density in g/cm^3. This volume is substituted into Eq 3 for calculation of the wear factor, *K.*

Coefficient of friction data were obtained with the thrust washer test apparatus. The test specimen is run in against the standard wear ring until 360° contact between the raise portion of the thrust washer and wear ring is achieved. Temperature of the test specimen is then allowed to stabilize at the test conditions (force of 280 kPa, or 40 psi; velocity of 0.25 m/s, or 50 ft/min). After thermal equilibrium occurs, the dynamic frictional torque generated is measured with the torque arm mounted on the antifriction bearing. The average of a minimum of five readings is taken.

In this study, the thrust washer testers were modified with thermal chambers that allowed independent control of the test temperature. Test specimens were allowed to come to equilibrium with the chamber temperature before the application of load and motion.

Material Selection

Identification of and formulations for the composites that were studied are found in Table 2. This material identification will be used throughout this discussion. The results of the wear and friction testing are presented in Tables 3, 4, 5, 6. Mechanical properties for the composites are presented in Table 7. The relative thermal stability of selected lubricant additives are shown in the thermogravimetric analysis (TGA) curves of Fig. 2.

Table 2 Composition of injection-moldable thermoplastic composites

Material	Base resin[(a)]	Reinforcement		Lubricant	
		Type	**%**	**Type**	**%**
LCL-4023	PEEK 150G	Carbon fiber	15	PTFE	10
LCL-4033 (150)	PEEK 150G	Carbon fiber	15	PTFE	15
LCL-4033 (450)	PEEK 450G	Carbon fiber	15	PTFE	15
PEEK PDX-88597	PEEK 450G	Carbon fiber	15	Intercalated graphite	5
PEEK PDX-88598	PEEK 450G	Carbon fiber	15	$SbSbS_4$	5
PEEK PDX-88273	PEEK 150G	Carbon fiber	15	CaF	10
PEEK PDX-88274	PEEK 150G	Carbon fiber	15	Graphite	10
PEEK PDX-88275	PEEK 150G	Carbon fiber	15	Graphite/PTFE	10/10
PEEK 150FT15	PEEK 150G	Carbon fiber	15	Talc	15
HTX-PDX-88599	HTX	Carbon fiber	30	. . .	. . .
PEK 220 GL30	PEK	Glass fiber	30	. . .	. . .
PEK 220 CA30	PEK	Carbon fiber	30	. . .	. . .
PEK PDX-88489	PEK	Carbon fiber	15	PTFE	15
PDX-88600	LCP (Xydar)	. . .	. . .	Graphite	10
PDX-88601	LCP (Xydar)	. . .	. . .	PTFE	10
VM-1	SRP	. . .	. . .	PTFE	. . .
VM-2	SRP	. . .	. . .	PTFE	. . .
JFL-4036	PES	Glass fiber	30	PTFE	15
JCL-4036	PES	Carbon fiber	30	PTFE	15
JCL-4063	PES	Carbon fiber	15	PTFE	30

(a) PEEK, polyether etherketone; HTX, high-temperature crystalline polymer; PEK, polyetherketone; LCP, liquid crystal polymer; SRP, self-reinforcing polymer; PES, polyether sulfone; PPS, polyphenylene sulfide.

(b) IPN, interpenetrating polymer network

Table 3 Wear and friction properties of PEEK composites

Material	Test parameters						Wear factor		Coefficient of friction	
	Temperature		**Load**		**Speed**					
	°C	**°F**	**kPa**	**psi**	**m/s**	**ft/min**	**Plastic**	**Steel**	**Static**	**Dynamic**
LCL-4023	23	73	280	40	0.25	50	13	0.3	0.08	0.17
	150	300	280	40	0.25	50	34	0.1	0.07	0.15
	260	500	280	40	0.25	50	120	0.03	0.08	0.17
LCL-4033 (150)	23	73	280	40	0.25	50	20	1.8	0.09	0.15
			35	5	2.0	400	5	1.0	. . .	. . .
	260	500	280	40	0.25	50	50	1.0	0.17	0.23
			35	5	2.0	400	30	1.0	. . .	. . .
	300	575	280	40	0.25	50	65	1.0	0.23	0.30
			35	5	2.0	400	45	1.0	Specimen melted	
LCL-4033 (450)	23	73	280	40	0.25	50	10	0.5	0.09	0.12
			35	5	2.0	400	15	0.5	. . .	. . .
	260	500	280	40	0.25	50	97	0.5	0.08	0.22
			35	5	2.0	400	13	0.5	. . .	. . .
L-1000	23	73	280	40	0.25	50	200	. . .	0.20	0.25
LL-4040	23	73	280	40	0.25	50	130	. . .	0.19	0.23
LC-1006	23	73	280	40	0.25	50	60	. . .	0.19	0.13
	95	200	280	40	0.25	50	60	. . .	0.11	0.06
	150	300	280	40	0.25	50	60	. . .	0.15	0.14
	205	400	280	40	0.25	50	120	. . .	0.12	0.09
	260	500	280	40	0.25	50	170	. . .	0.15	0.14
PEEK 150HT15	23	73	280	40	0.25	50	160	0.3	0.06	0.18
			35	5	2.0	400	125	0.2	0.35	0.48
	260	500	280	40	0.25	50	75	1.0	0.04	0.05
			35	5	2.0	400	60	0.7	0.49	0.61
PEEK PDX-88597	23	73	280	40	0.25	50	25	0.0	0.08	0.15

	150	300	280	40	0.25	50	27	0.3	0.11	0.14
	260	500	280	40	0.25	50	28	0.4	0.11	0.17
PEEK PDX-88598	23	73	280	40	0.25	50	13	1.0	0.11	0.11
	150	300	280	40	0.25	50	78	1.0	0.14	0.14
	260	500	280	40	0.25	50	120	1.0	0.15	0.15
PEEK PDX-88273	23	73	280	40	0.25	50	14	0.9	0.13	0.18
			35	5	2.0	400	4	0.3	0.06	0.09
	260	500	280	40	0.25	50	155	1.2	0.08	0.13
			35	5	2.0	400	30	0.3	0.40	0.51
PEEK PDX-88274	23	73	350	50	0.25	50	18	1.3	0.05	0.15
			35	5	2.0	400	7	0.6	0.08	0.09
	260	500	280	40	0.25	50	70	0.2	0.10	0.13
			35	5	2.0	400	15	1.7	0.15	0.11
PEEK PDX-88275	23	73	280	40	0.25	50	10	0.4	0.07	0.20
			35	5	2.0	400	6	0.4	0.17	0.19
	260	500	280	40	0.25	50	40	0.6	0.13	0.22

Table 4 Wear and friction properties of PEK, HTX, and SRP composites

Material	Test parameters						Wear factor		Coefficient of friction	
	Temperature		**Load**		**Speed**					
	°C	**°F**	**kPa**	**psi**	**m/s**	**ft/min**	**Plastic**	**Steel**	**Static**	**Dynamic**
HTX PDX-88599	23	73	280	40	0.25	50	56	8.9	0.07	0.23
			35	5	2.0	400	65	3.7	0.11	0.09
	260	500	280	40	0.25	50	60	0.6	0.19	0.29
			35	5	2.0	400	84	0.2	0.22	0.17
PEK 220 GL30	23	73	280	40	0.25	50	109	11.1	. . .	0.36
PEK 220 CA30	23	73	280	40	0.25	50	58	0.5	0.14	0.14
	260	500	280	40	0.25	50	66	0.2	0.24	0.27
PEK PDX-88489	23	73	280	40	0.25	50	18	1.4	0.21	0.20
	260	500	280	40	0.25	50	86	0.1	0.22	0.19
PDX-88600	23	73	280	40	0.25	50	92	0.1	0.09	0.16
	150	300	280	40	0.25	50	1140	0.1	0.10	0.15
PDX-88601	23	73	280	40	0.25	50	30	0.6	0.13	0.17
	150	300	280	40	0.25	50	46	0.2	0.09	0.10
VM-1	23	73	280	40	0.25	50	4	0.4	0.04	0.12
	150	300	280	40	0.25	50	18	2.8	0.14	0.19
VM-2	23	73	280	40	0.25	50	3	0.3	0.08	0.14

Table 5 Wear and friction properties of PPS composites

Material	Test parameters						Wear factor		Coefficient of friction	
	Temperature		**Load**		**Speed**					
	°C	**°F**	**kPa**	**psi**	**m/s**	**ft/min**	**Plastic**	**Steel**	**Static**	**Dynamic**
Lubricomp O-BG(a)	23	73	280	40	0.25	50	35	6.0	0.07	0.18
			35	5	2.0	400	23	0.9	0.09	0.10
	150	300	280	40	0.25	50	31	0.2	0.11	0.17
			35	5	2.0	400	21	0.9	0.11	0.24
	175	350	280	40	0.25	50	54	2.1	0.09	0.18
Lubricomp O-BG(b)	175	350	280	40	0.25	50	51	2.0	0.07	0.10
Lubricomp O-BG(a)	205	400	280	40	0.25	50	Specimen melted		Specimen melted	
Lubricomp O-BG(b)	205	400	280	40	0.25	50	36	1.1	0.09	0.20
			35	5	2.0	400	Specimen melted		Specimen melted	
Lubricomp O-BG(a)	260	500	280	40	0.25	50	Specimen melted		Specimen melted	
			35	5	2.0	400	390	0.3	0.78	0.97
Lubricomp O-BG(b)	260	500	35	5	2.0	400	Specimen		Specimen	

(a) Low mold temperature.

(b) High mold temperature.

Table 6 Wear and friction properties of PES composites

Specimens tested at a load of 280 kPa (40 psi) and a speed of 0.25 m/s (50 ft/min)

Material	Test temperature		Wear factor		Coefficient of friction	
	°C	**°F**	**Plastic**	**Steel**	**Static**	**Dynamic**
JFL-4036	23	73	60	3.0	0.16	0.20
	95	200	120	2.5	0.27	0.30
JCL-4036	23	73	40	2.0	0.13	0.17
	95	200	97	1.5	0.24	0.29

Table 7 Mechanical properties of selected thermoplastic composites

Material	Specific gravity (ASTM D 792)	Tensile strength (ASTM D 638)		Tensile elongation (ASTM D 638), %	Flexural strength (ASTM D 790)		Flexural modulus (ASTM D 790)		Izod impact (ASTM D 256), ft · lbf/in.		Deflection temperature at 1.72 MPa (254 psi) (ASTM D 648)	
		MPa	**ksi**		**MPa**	**ksi**	**GPa**	**10^6 psi**	**Notched**	**Unnotched**	**°C**	**°F**
LCL-4023	1.37	165	24.0	3.5	262	38.0	12.07	1.750	1.5	9	315	600
LCL-4033 (150)	1.41	163	23.6	7.0	243	35.2	12.51	1.815	1.8	12	315	600
LCL-4033 (450)	1.41	177	25.6	4.5	256	37.1	12.39	1.797	1.5	11	315	600
PEEK PDX-88597	1.35	153	22.2	3.1	238	34.5	11.86	1.720	1.1	10	315	600
PEEK PDX-88598	1.36	176	25.5	3.5	218	31.6	10.55	1.530	1.2	9	315	600
PEEK PDX-88273	1.41	157	23.7	3.8	265	38.5	11.81	1.713	1.3	11	315	600
PEEK PDX-88274	1.41	172	24.9	3.7	282	40.9	13.38	1.940	1.3	11	315	600
PEEK PDX-88275	1.46	161	23.4	3.8	253	36.7	12.48	1.810	1.3	10	315	600
PEEK 150HT15	1.39	88	12.7	6.5	142	20.6	4.82	0.700	1.1	14	315	600
HTX PDX-88599	. . .	183	26.6	3.8	247	35.8	15.07	2.185	1.6	10	324	615
PEK 220 GL30	1.43	170	24.7	4.0	234	34.0	8.96	1.300	1.7	. . .	307	676
PEK 220 CA30	1.40	238	34.5	3.0	359	52.0	18.44	2.675	1.5	16	360	680
PEK PDX-88489	1.44	192	27.9	5.4	285	41.4	12.27	1.780	1.5	12	359	678
PDX-88600	. . .	125	18.1	4.3	106	15.4	11.51	1.670	1.1	8	. . .	. . .
PDX-88601	. . .	114	16.5	5.1	90	13.1	9.31	1.350	1.1	4	. . .	. . .
VM-1	. . .	100	14.5	4.9	. . .	. . .	6.55	0.950	. . .	. . .	. . .	. . .
VM-2	. . .	55	8.0	3.6	. . .	. . .	4.72	0.685	. . .	. . .	. . .	. . .
OFL-4036	1.70	124	18.0	4.0	. . .	. . .	7.58	1.100	1.5	. . .	210	410
OCL-4036	1.57	162	23.5	3.0	. . .	. . .	13.10	1.900	1.0	. . .	210	410
Lubricomp O-BG	. . .	128	18.6	3.8	194	28.1	16.20	2.350	1.6	7	260	500

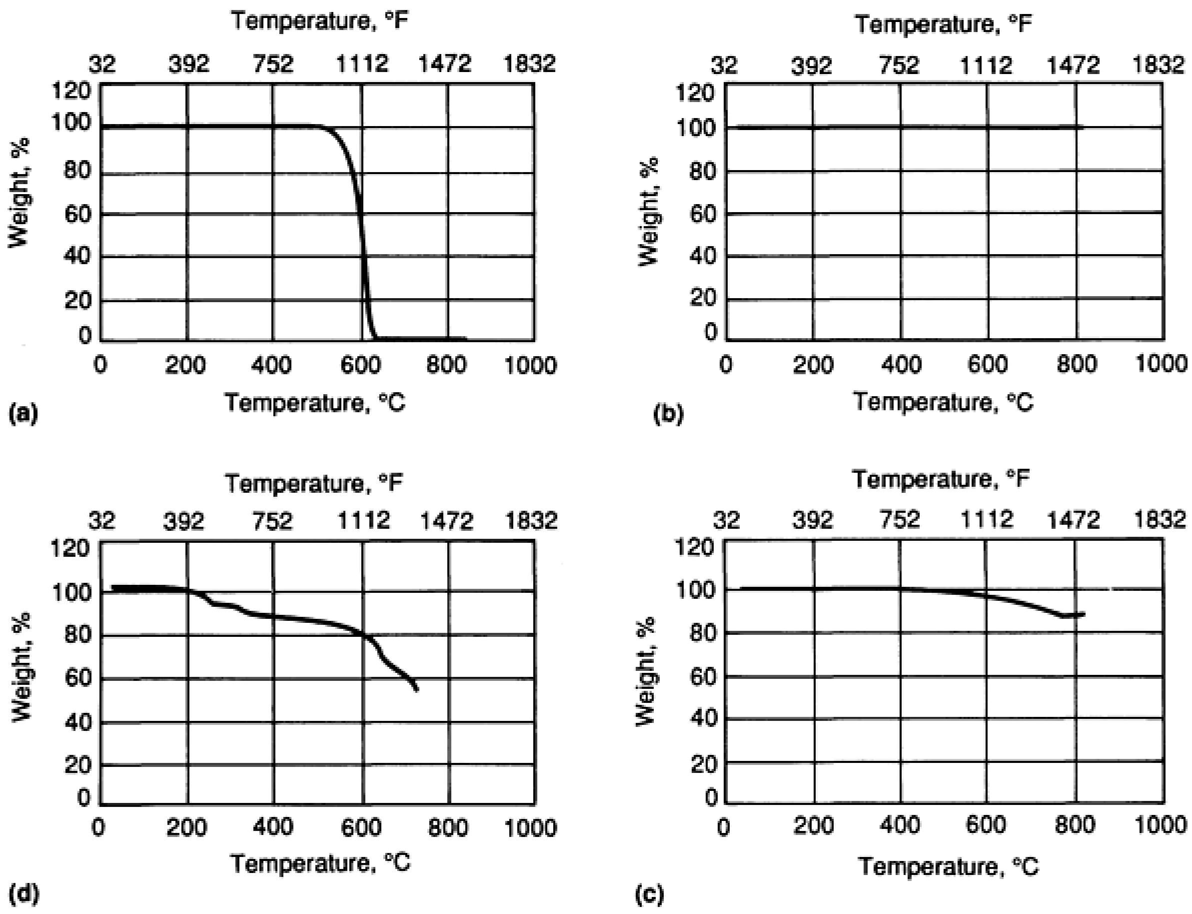


Fig. 2 Thermogravimetric analyses of lubricant additives. (a) PTFE. (b) Calcium fluoride. (c) Intercalated graphite. (d) $SbSbS_4$

Wear and friction data for polyether etherketone (PEEK) resin and composites are presented in Table 3. While all additives in this study reduced the wear factor, it is important to note that the addition of fibrous reinforcement (for example, carbon fibers) is required to develop minimum wear at elevated temperatures. Wear factor and both coefficients of friction increase with increasing temperature in the PEEK composites evaluated.

Effect of Lubricant on Wear Factor. As expected, compounds using high-molecular-weight PEEK resins (for example, PEEK 450G) had lower wear factors than identical compounds based on lower molecular weight resins (for example, PEEK 150G) at room temperature. However, at 260 °C (500 °F) the reverse is true; that is, PEEK 150G materials has lower wear factors (Fig. 3). This is due to the improved bonding of the fiber at the polymer interface.

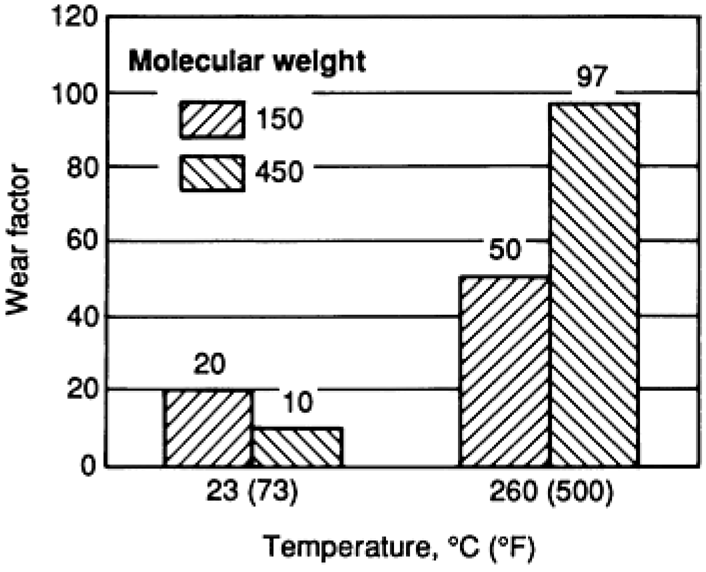


Fig. 3 Effect of molecular weight on the wear factors of carbon fiber/PTFE/PEEK composites

PTFE and graphite powder are often used as internal lubricants for thermoplastic composites. They have been found to be effective lubricants in PEEK. At 23 °C (73 °F), graphite-powder-lubricated carbon-fiber-reinforced PEEK will demonstrate wear and frictional behavior similar to that of the analogous PTFE-lubricated composite. However, at 260 °C (500 °F), the graphite-lubricated composite demonstrates 40% more wear than the PTFE-lubricated composite (Fig. 4).

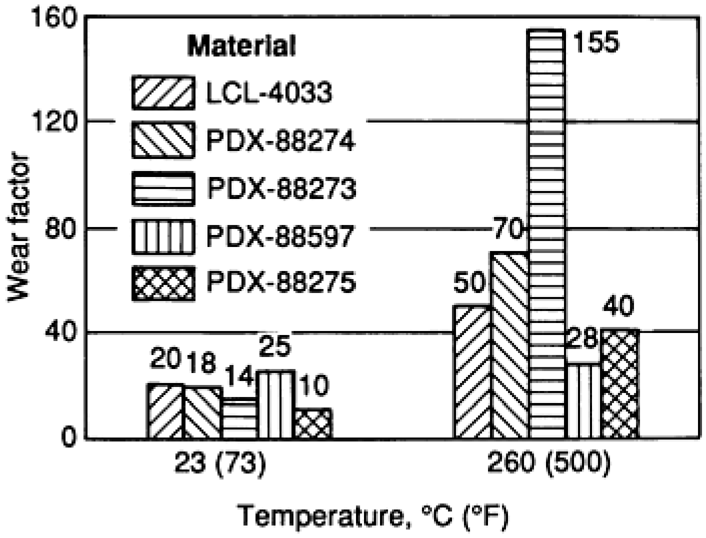


Fig. 4 Effect of lubricant on the wear factors of carbon fiber/PEEK composites

Graphite powder was evaluated in combination with PTFE lubricant in carbon-fiber-reinforced PEEK (PEEK PDX-88275). This lubricant package demonstrated lower wear factors than either lubricant individually at 23 °C (73 °F) and 260 °C (500 °F) (Fig. 4).

Calcium fluoride demonstrated very good lubricating behavior in carbon-fiber-reinforced PEEK at room temperature. However, calcium fluoride is not an effective lubricant at 260 °C 500 °F) (Fig. 4).

Intercalated graphite demonstrated low wear factors when used as an internal lubricant in carbon-fiber-reinforced PEEK at all temperatures. The wear factor for this composite (PDX-88597 at 260 °C (500 °F) was identical to the wear factor at room temperature. At 260 °C (500 °F), the wear factor for the composite was one-fourth that of the analogous PTFE-lubricated composite (LCL-4033) (Fig. 4).

While $SbSbS_4$ is a very effective lubricant in carbon-fiber-reinforced PEEK at room temperature, it loses this effectiveness at 260 °C (500 °F) when compared to the analogous PTFE-lubricated composite (PEEK PDX-88598) (Fig. 5).

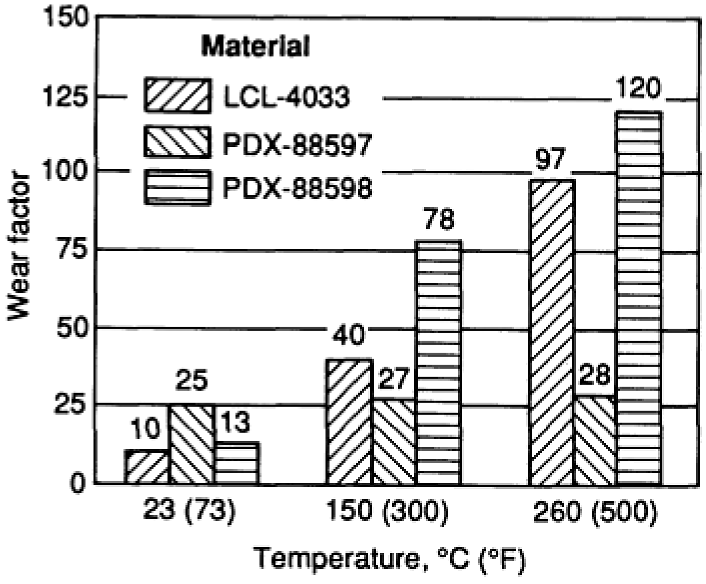


Fig. 5 Effect of lubricant on the wear factors of carbon fiber/PEEK composites

The wear factors of PEEK composites are affected by the load, *P,* and speed components, *V,* independent of total *PV.* All composites demonstrated lower wear factors at low pressures (35 kPa, or 5 psi) and high speeds (2.0 m/s, or 400 ft/min) than at high pressures (280 kPa, or 40 psi) and intermediate speeds (0.25 m/s, or 50 ft/min). This was demonstrated at all temperatures (Fig. 6).

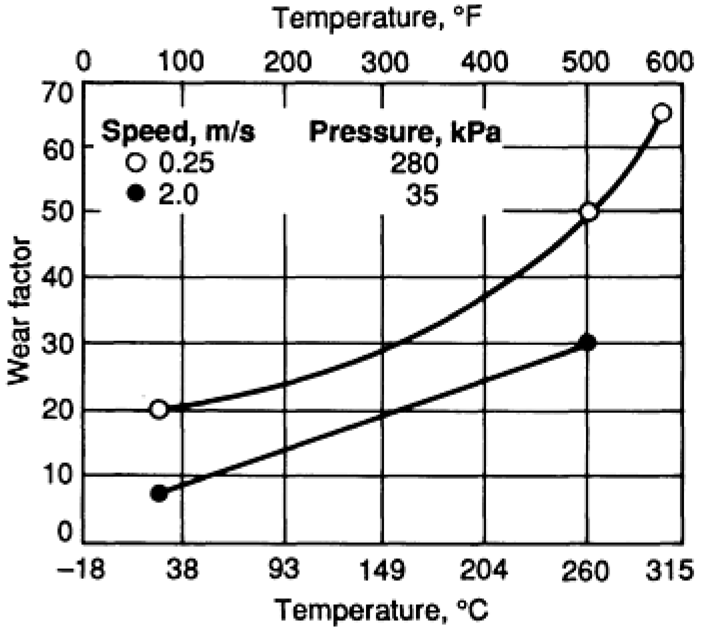


Fig. 6 Plot of wear factor versus test temperature as a function of speed and load for LCL-4033 composite

Carbon-fiber-reinforced PEEK, polyetherketone (PEK), and high-temperature-crystalline (HTX) composites demonstrated similar wear factors at room temperature. The wear factor of carbon-fiber-reinforced PEEK increased threefold when tested at 260 °C (500 °F). (The wear factor of carbon-fiber-reinforced PEEK at 260 °C, or 500 °F, is less than the wear factor of unfilled nylon 6/6 at room temperature.) The wear rates for carbon-fiber-reinforced PEK and HTX were almost the same at 260 °C (500 °F) as at room temperature (Fig. 7).

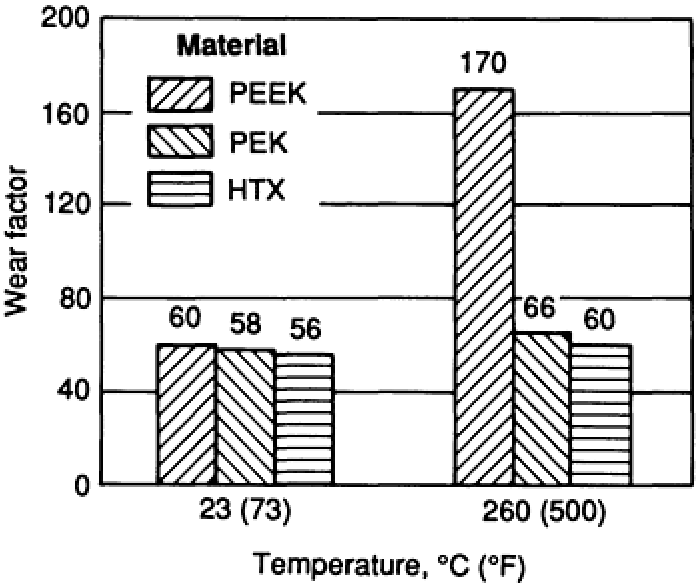


Fig. 7 Effect of temperature on the wear factors of 30% carbon-fiber-reinforced PEEK, PEK, and HTX composites

Polyphenylene sulfide (PPS) is similar to PEEK composites. Lubricomp O-BG demonstrated lower wear factors at 35 kPa (5 psi) and 2.0 m/s (400 ft/min) than at 280 kPa (40 psi) and 0.25 m/s (50 ft/min). As in previous studies for PPS compounds (Ref 1), Lubricomp O-BG displays a maximum service temperature of 205 °C (400 °F) as a bearing material (Fig. 8).

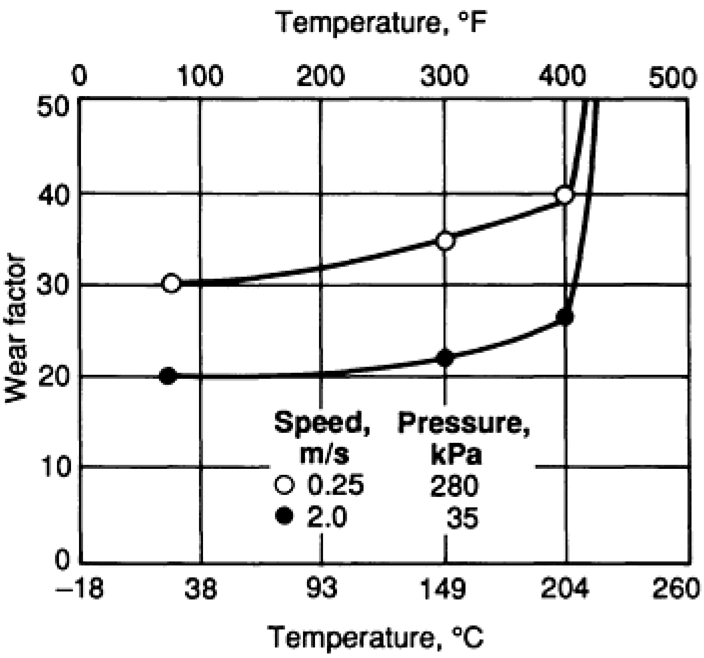


Fig. 8 Plot of wear factor versus test temperature as a function of speed and load for Lubricomp O-BG composite

The wear performance of two different liquid crystal polymers (LCP) was determined in the following study, PTFE reduces the wear factor to a greater extent than graphite at both 23 °C (73 °F) and 150 °C (300 °F). The experimental self-reinforcing polymer (SRP), when lubricated with PTFE, demonstrates significantly lower wear factors and coefficients of friction at room temperature than the liquid crystal polymer Xydar. At 150 °C (300 °F) the SRP/PTFE also demonstrates lower wear and slightly higher coefficients of friction than the LCP/PTFE.

Effect of Lubricants on Friction. There is a distinct trend for compounds based on high-molecular-weight PEEK resins to demonstrate lower static and dynamic coefficients of friction at both room temperature and at 260 °C (500 °F) than for those based on low-molecular-weight PEEK. The greatest difference is found in the static coefficient of friction at 260 °C (500 °F) for the 15% PTFE-lubricated, 15% carbon-fiber-reinforced PEEK composites (Fig. 9).

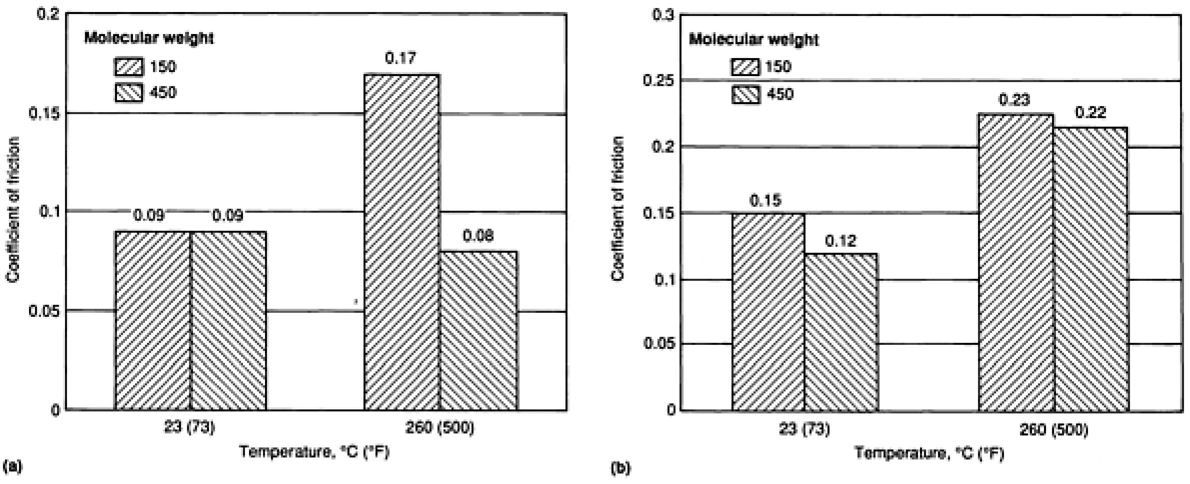


Fig. 9 Effect of molecular weight and test temperature on the coefficients of friction of carbon fiber/PTFE/PEEK composites. (a) Static coefficient of friction. (b) Dynamic coefficient of friction

There are no clear-cut trends for the effects of lubricants on the coefficients of friction. PTFE lubricants demonstrate the lowest static coefficients of friction at all temperatures when compared to metallic lubricants (intercalated graphite, $SbSbS_4$ and CaF_2). However, graphite powder can offer a lower static coefficient of friction than PTFE (Fig. 10 and 11).

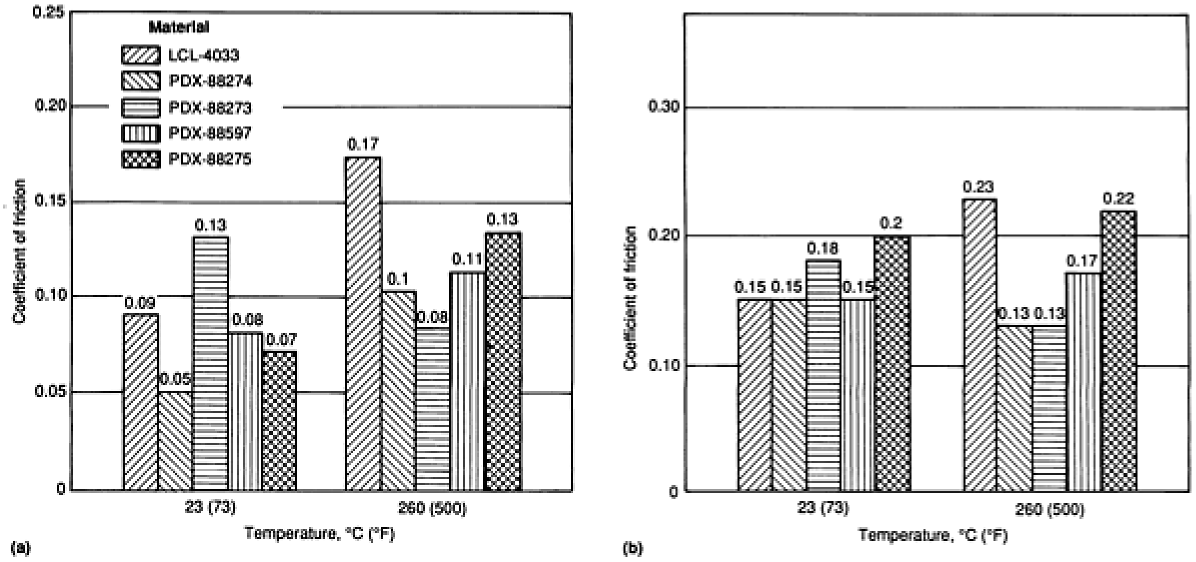


Fig. 10 Effect of lubricant and test temperature on the coefficients of friction of carbon fiber/PEEK composites. (a) Static coefficient of friction. (b) Dynamic coefficient of friction

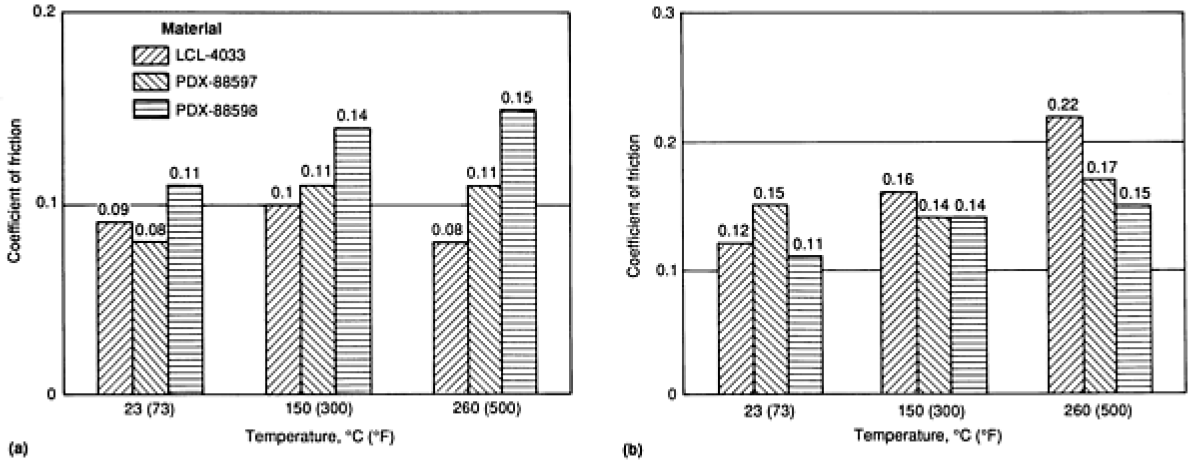


Fig. 11 Effect of lubricant and test temperature on the coefficients of friction of carbon fiber/PEEK composites. (a) Static coefficient of friction. (b) Dynamic coefficient of friction

Coefficients of friction are expected to decrease with increasing thermal stability of the resin (for example, melt point, T_g, or continuous-use temperature). This trend is *not* well demonstrated by 30% carbon-fiber-reinforced PEEK, PEK, or HTX (Fig. 12).

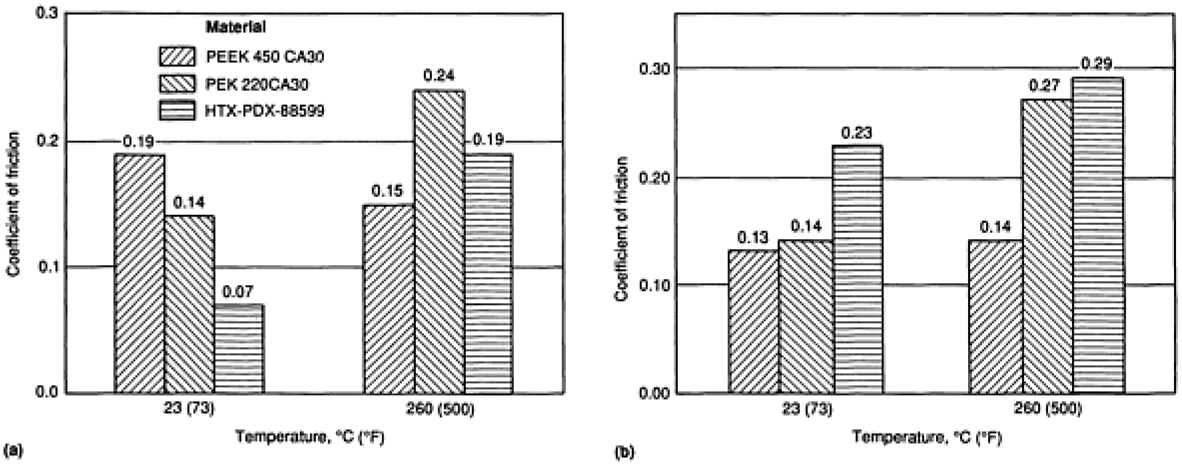


Fig. 12 Effect of test temperature on the coefficients of friction of 30% carbon-fiber-reinforced PEEK, PEK, and HTX composites. (a) Static coefficient of friction. (b) Dynamic coefficient of friction

The coefficient of friction increased for the PTFE-lubricated, carbon-fiber-reinforced PEEK composite with increasing temperature (Fig. 13).

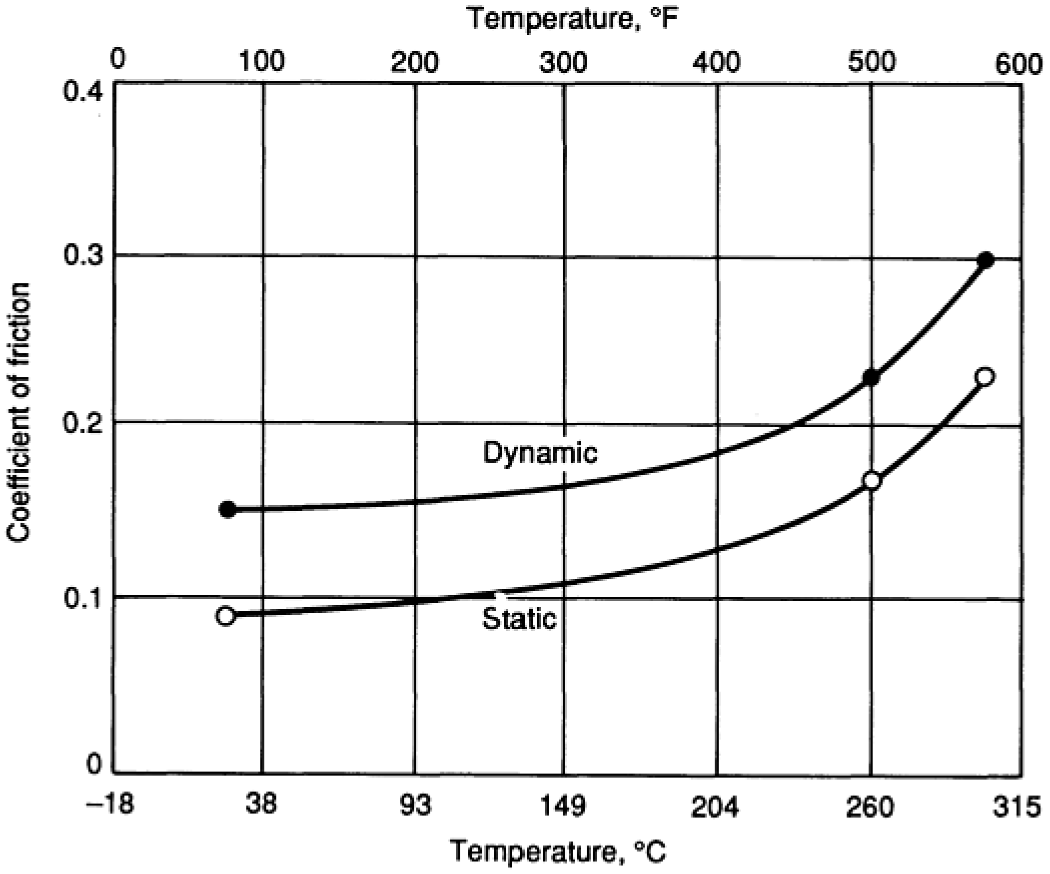


Fig. 13 Plot of coefficient of friction versus test temperature for LCL-4033

The reinforced, lubricated PPS composite demonstrates a clear trend for the dynamic coefficient of friction to increase with temperature. The static coefficient of friction reaches a maximum at approximately 150 °C (300 °F) and then rapidly diminishes with further increases in temperature (Fig. 14).

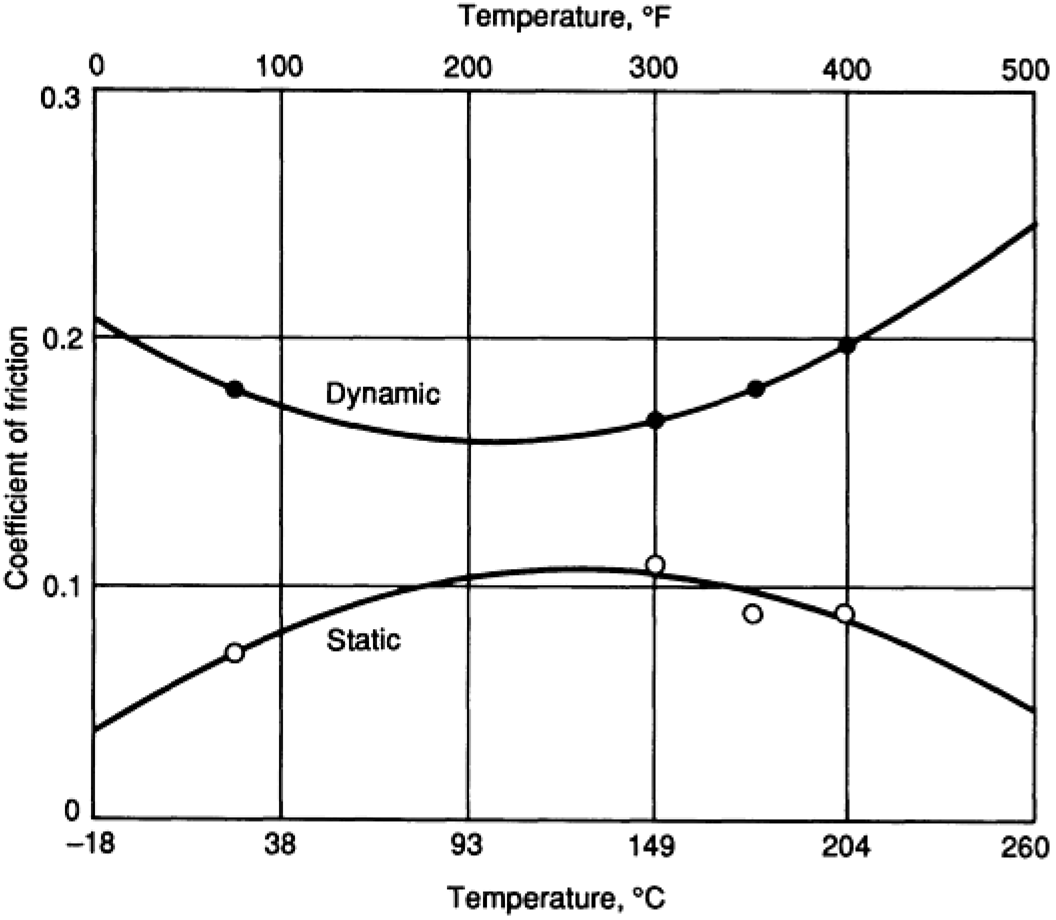


Fig. 14 Plot of coefficient of friction versus test temperature for Lubricomp O-BG

Conclusions. The results of the tests conducted on the composites listed in Table 2 can be summarized as follows:

- PEEK composites formulated with lower molecular weight resin demonstrated lower wear factors than analogs using higher molecular weight resin at elevated temperatures. This is probably due to superior wet-out of the resin and fiber in the low-molecular-weight composites
- Intercalated graphite is a very effective (and expensive) high-temperature lubricant for carbon-fiber-reinforced high-molecular-weight PEEK
- For service to 205 °C (400 °F), Lubricomp O-BG (PPS) and LCL-4033 (PEEK) composites are the materials of choice for wear and friction applications
- For service at 260 °C (500 °F), lubricated PEEK composites are the materials of choice for wear and friction applications
- For service at temperatures in excess of 260 °C (500 °F), composites based on PEK and HTX will be required for wear and friction applications

References

1. M.P. Wolverton, J.E. Theberge, and K.L. McCadden, How Plastics Wear at High Temperatures, *Mach. Des.,* Vol 55 (No. 3), p 111
2. K. Friedrick, "Friction and Wear of Composites," Composites, Workshop, Center for Composite Materials, University of Delaware, 16 May 1986
3. J.F. Archard and W. Hirst, Wear of Metals under Unlubricated Conditions, *Proc. R. Soc. (London) A,* Aug 1956

Thermal Spray Coatings

Burton A. Kushner and Edward R. Novinski, Perkin-Elmer Corporation, Metco Division

Introduction

THE THERMAL SPRAY COATING PROCESS is by far the most versatile modern surfacing method with regard to economics, range of materials, and scope of applications. The thermal spray process permits rapid application of high-performance materials in thicknesses from a few mils to more than 25 mm (1 in.) on parts of a variety of sizes and geometries. Thermal spraying requires minimal base-metal preparation, can be applied in the field, and is a low-temperature (>95 °C, or >200 °F) method compared with techniques such as weld overlay. Typical part configurations include piston rings, journals, conveyors, shifter forks, extrusion dies, transformer cases, ship hulls, ship tanker compartments, and suspension bridges.

Thermal spraying reduces wear and corrosion and greatly prolongs part service life by allowing use of a high-performance coating material over a low-cost base metal. Application areas of the process can be categorized as (Ref 1):

- Wear resistance
- Oxidation resistance
- Corrosion resistance
- Restoration of dimension
- Abradable clearance control
- Thermal barriers
- Electrical conductivity or resistivity
- Biomedical

More than 200 coating materials with different characteristics of toughness, coefficient of friction, hardness, and other properties are available. These materials can be grouped as follows:

- Pure metals
- Metal alloys
- Cermets
- Ceramics
- Carbides
- Polymers
- Special composite materials

All thermal spray processes rely on three basic operational mechanisms:

- Heating a coating material in either wire or powder form to a molten or plastic state
- Propulsion of particles of the heated material
- Impact of the material onto a workpiece whereby the particles rapidly solidify and adhere both to one another and to the substrate to form a dense, functional, protective coating

The particles bond to the substrate mechanically and, in some cases, metallurgically. Particle velocity, substrate roughness, particle size, material chemistry, particle temperature, and substrate temperature influence the bond strength of the coating material. The process was originally referred to as flame spraying, metal spraying, flame plating, or metallizing when it was limited to the oxygen-fuel (oxyfuel) wire spray method.

Thermal Spray Processes

Currently, five different commercially available thermal spray methods are in use:

- Oxyfuel wire (OFW) spray
- Electric arc wire (EAW) spray
- Oxyfuel powder (OFP) spray
- Plasma arc (PA) powder spray
- High-velocity oxyfuel (HVOF) powder spray

Selection of the appropriate thermal spray method is typically determined by:

- Desired coating material
- Coating performance requirements
- Economics
- Part size and portability

The oxyfuel wire spray process (also called wire flame spraying or the combustion wire process) is the oldest of the thermal spray coating methods and among the lowest in capital investment. The process utilizes an oxygen-fuel gas flame as a heating source and coating material in wire form. Solid rod feed stock has also been used. During operation, the wire is drawn into the flame by drive rolls that are powered by an adjustable air turbine or electric motor (Fig. 1). The tip of the wire is melted as it enters the flame and is atomized into particles by a surrounding jet of compressed air and propelled to the workpiece.

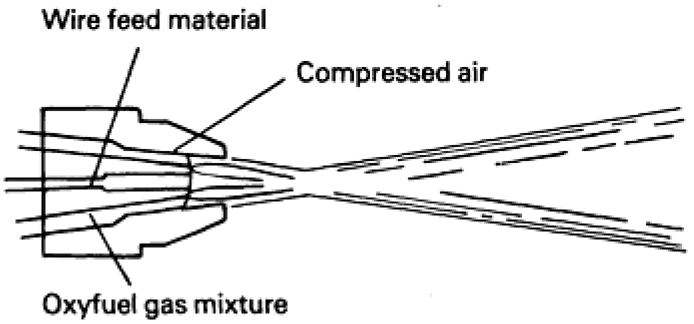


Fig. 1 Atomization of wire feedstock from the nozzle of an OFW spray gun

Spray rates for this process range from 2.3 to 55 kg/h (5 to 120 lb/h) and are dictated by the melting point of the material and the choice of fuel gas. The wire spray gun is most commonly used as a hand-held device for on-site application, although an electric-motor-driven gun is recommended for fixed-mounted use in high-volume, repetitive production work.

The OFW process is widely used for corrosion protection of large outdoor structures, such as bridges and storage tanks, and for restoration of dimension to worn machinery components. It is a good choice for all-purpose spraying. Coatings can be applied rapidly and at low cost, and a wide variety of metal coating materials are available. Typical spray materials include austenitic and martensitic stainless steels, nickel aluminide, nickel chromium alloy, bronze, Monel, babbitt, aluminum, zinc and molybdenum.

Electric arc wire spraying also applies coatings of selected metals in wire form. Push-pull motors feed two electrically charged wires through the arc gun to contact tips at the gun head (Fig. 2). An arc is created that metals the wires at temperatures above 5500 °C (10,000 °F). Compressed air atomizes the molten metal and projects it onto a prepared surface.

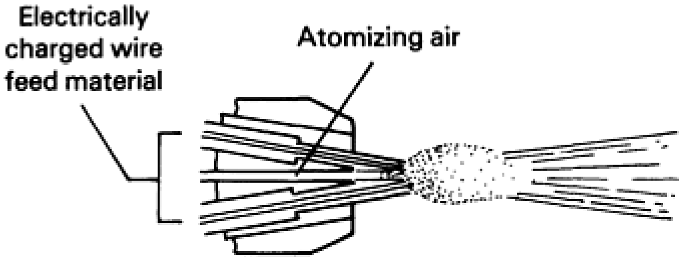


Fig. 2 Spray pattern generated by two electrically charged wires melted at the nozzle of an EAW spray gun

The EAW process is excellent for applications that require heavy coating buildup or that have large surfaces to be sprayed. The arc system can produce a spray pattern ranging from 50 to 300 mm (2 to 12 in.) and can spray at high speeds. It has built-in flexibility, allowing coating characteristics, such as hardness or surface texture, to be tailored to specific applications.

The EAW method is characterized by strong coating adhesion because of the high particle temperatures produced. Because the process uses only electricity and compressed air, it allows equipment to be moved relatively easily from one installation to another, and eliminates the need to stock oxygen and fuel gas supplies. Materials applied by the EAW method are similar to those used in the OFW process.

The oxyfuel powder spray method extends the range of available coatings and subsequent applications to include ceramics, cermets, carbides, and fusible hardfacing coatings. Using either gravity flow or pressurized feed, powder is fed into the gun and carried to the gun nozzle (Fig. 3), where it is melted and projected by the gas stream onto a prepared surface. For general-purpose spraying, the gravity-flow system is used. When exacting coating consistency and/or high spray rates are desired, the pressurized feed system is used. Oxyfuel powder guns are the lowest cost thermal spray equipment and are easiest to set up and change coating materials. The OFP method finds widest use in short-run machinery maintenance work and in the production spraying of abradable clearance-control seals for gas turbine engines.

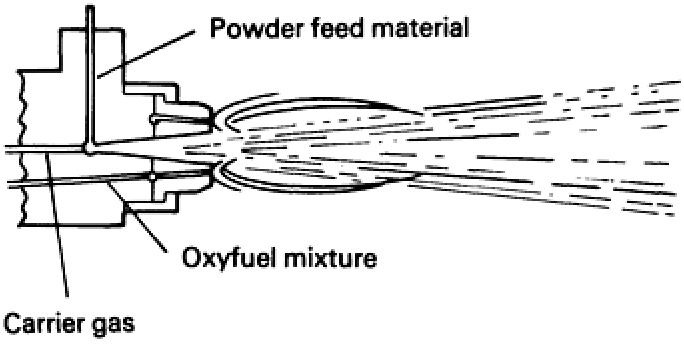


Fig. 3 Cross-sectional view of an OFP spray system showing powder feed material being transported by the carrier gas and then melted by the oxy-fuel mixture

Plasma arc powder is one of the most sophisticated and versatile thermal spray methods. Temperatures that can be obtained with commercial plasma equipment have been calculated to be greater than 11,000 °C (20,000 °F) and are far above the melting or even the vaporization point of any known material. Decomposition of materials during spraying is minimized because of the high gas velocities produced by the plasma, resulting in extremely short residence time in the thermal environment. The plasma process also provides a controlled atmosphere for melting and transport of the coating material, thus minimizing oxidation, and the high gas velocities produce coatings of high density.

The plasma gun operates on the principle of raising the energy state of a gas by passing it through an electric arc. The release of energy in returning the gas to its ground state results in exceedingly high temperatures. A gas such as nitrogen

or argon enters a direct-current arc between a tungsten cathode and a copper anode that make up the nozzle (Fig. 4). Both components are cooled by a constant flow of water through internal passages. Here the plasma gas first dissociates (in the case of nitrogen, into two atoms), followed by ionization that releases free electrons. The electrons recombine outside the electric arc, and energy is released as heat and light. In addition, frequent collisions transfer energy from the electrons to the positive ions, accelerating them until the plasma reaches a state of equilibrium. The result is a thermal plasma, in which the energy of the electrons has been turned into enthalpy, or heat content (Ref 2). At this point, powdered coating material suspended in a gas is injected into the plasma and is subsequently melted and propelled at high velocity to the workpiece. In practice, a small amount of a secondary gas, such s hydrogen or helium, is mixed with the primary plasma gas to increase operating voltage and thermal energy.

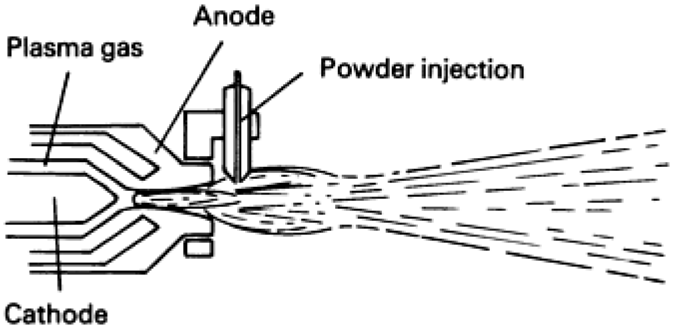


Fig. 4 Schematic of a plasma arc powder spray system showing routing of plasma gas and powder material at the output nozzle

The high temperatures and high gas velocities produced by the plasma process result in coatings that are superior in mechanical and metallurgical properties to low-velocity OFW or OFP coatings. The plasma process is particularly efficient for spraying high-quality coatings of ceramic materials, such as zirconium oxide for turbine engine combustors and chromium oxide for printing rolls. The plasma process is also readily field-portable and is thus used for large on-site applications such as power utility plant boiler tubes.

Current plasma spray technology permits fully automatic start/stop operation and closed-loop computer control for power level, plasma gas flow, and powder feed rate. System problems can be diagnosed via computer modem to the equipment manufacturer, and system performance can be documented and stored on a digital recorder or data logger. A variation of the plasma spray process is to conduct spraying within a vacuum or low-pressure chamber. Although this significantly adds to cost, the quality of metallic coatings is improved by minimizing oxides within the coating, reducing porosity, and increasing coating adhesion.

With any thermal spray powder method, the degree to which a given flame effectively melts and accelerates the powder depends on the type of coating material and the size and shape of the particles. Each particular coating material and gun combination has an optimum particle size. Particles much smaller than ideal will overheat and vaporize; much larger particles will not melt and may fall from the flame or rebound from the target (Ref 2).

The high velocity oxyfuel powder spray process (also known as the hypervelocity oxy-fuel powder spray process, the oxyfuel detonation [OFD] process, and the D-gun process) represents the state of the art for thermal spray metallic coatings. The HVOF process uses extremely high kinetic energy and controlled thermal energy output to produce very-low-porosity coatings that exhibit high bond strength, fine as-sprayed surface finish, and low residual stresses.

The HVOF process with an oxygen-fuel mixture consisting of oxygen and either acetylene, propylene, propane, or hydrogen fuel gas, depending on coating requirements. The fuel gas flows through a siphon system, where it is thoroughly mixed with oxygen (Fig. 5a). In one design, the mixed gases are ejected from the gun nozzle and are ignited. The high-velocity gases produce uniquely characteristic multiple shock diamond patterns, which are visible in the flame (Fig. 5b). Combustion temperatures approach 2750 °C (5000 °F) and form a circular flame configuration. Powder is injected into the flame axially to provide uniform heating, and powder particles are accelerated by the high-velocity gases, which typically approach a speed of 1350 m/s (4500 ft/s).

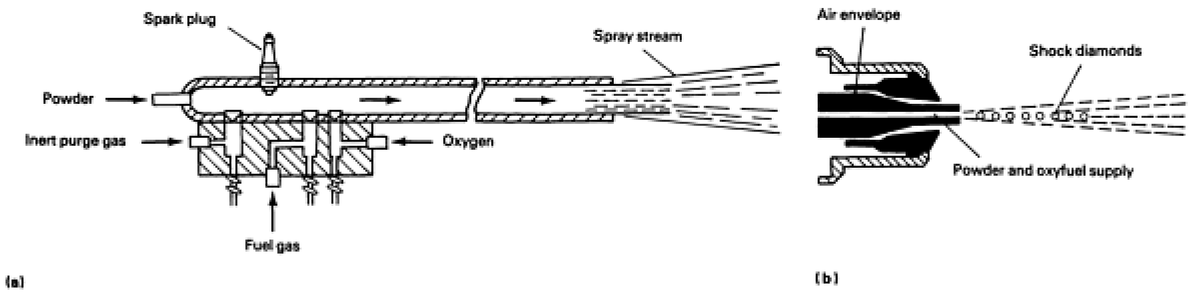


Fig. 5 Schematics showing input and output of an HVOF powder spray device. (a) Key components of an HVOF system. (b) Close-up view of HVOF spray system output

The low residual coating stress produced in the HVOF process allows significantly greater thickness capability than the plasma method, while providing lower porosity, lower oxide content, and higher coating adhesion. Coatings produced by the HVOF process also have much better machinability compared with other methods, and coating porosity has closely approached wrought materials, as verified by recent gas permeability testing. HVOF systems are available with closed-loop computer control and robotics capability.

Process Parameters

Several important factors must be considered when selecting an appropriate thermal spray method:

- Surface preparation
- Deposition rate
- Coating thickness limitations
- Bond coat materials
- Coating finishing method

Surface Preparation. Bonding of thermal spray coatings relies primarily on mechanical interlocking with the substrate material. Good surface preparation cannot be overemphasized; most coating failures can be traced to poor practice at this first step of the operation.

Nonporous surfaces should be cleaned to remove organic contamination by vapor degreasing or by washing with hot detergent solutions or steam cleaning. Porous castings may require heat treatment to pyrolize contamination at approximately 200 to 300 °C (400 to 600 °F). Cleaned parts should be uniformly abrasive blasted to achieve a white metal condition and a minimum of 6 μm (250 μin.) R_a (arithmetic average surface roughness) finish using either aluminum oxide or chilled, angular iron steel grit. The grit and air supply must be oil-free so that the cleaned surfaces are not recontaminated. Blasted parts should be handled with clean gloves and protected from shop soil until the coating operation. Spraying should be done within 2 h of blasting. Parts should be reblasted if this time interval is exceeded. Machined undercuts should have at least a 45° taper and must be cleaned and grit blasted as described above.

Deposition Rate. The coating deposition rate is limited by the method used and the melting point of the coating material. Other considerations include the deposit efficiency and target efficiency. Deposit efficiency is the quantity of coating material deposited relative to the quantity being sprayed. Target efficiency relates to the area of the spray pattern relative to the size of the part or target. Both factors influence cost.

Coating Thickness Limitations. All thermal spray coatings exhibit a degree of internal stress as a result of shrinkage from a molten state to a solid state. These stresses accumulate as the coating thickness increases and result in a shear force at the substrate interface. Ductile coating materials tend to exhibit low stress; the opposite is true for hard coating materials, such as carbides or ceramics. Also, very porous coatings exhibit lower stress than denser coatings. When the internal stress exceeds the adhesion, the coating can delaminate from the substrate or crack. Equipment manufacturers usually provide the practical thickness limit for each spray material.

Bond Coat Materials. Bond coatings form a metallurgical bond with the substrate. They include materials that react with the substrate exothermically, such as nickel-aluminum alloys, and materials that melt at high temperatures, such as molybdenum.

These materials are frequently applied as a thin coating under a poor-bonding or high-stress topcoat to enhance the adhesion provided by grit blasting alone. Bond coat materials are typically low in internal stress, and single coats are sufficient if the physical properties meet the design specifications. Selection of a particular bond coating depends on its compatibility with the spray method being used in terms of parameters such as bond strength, thermal expansion properties related to the base metal, corrosion resistance, and oxidation resistance.

Coating Finishing Method. Thermal spray coatings are finished to dimension either by single-point machining or by grinding. Hard surfacing materials such as ceramics and carbides are restricted to grind finishing, typically with diamond wheels. Thermal spray coatings are sufficiently different from the same material in wrought form that different grinding wheel and finishing tool recommendations are almost always required. Therefore, tools and wheels should not be chosen based on experience with the parent material in wrought or cast form (Ref 3).

Most thermal spray coatings exhibit some degree of porosity. However, the HVOF method produces coatings closest in machining and grinding behavior to wrought materials. In general, thermal spray coatings can be machined to 1 to 2 μm (40 to 80 μin.) R_a and ground to 0.25 to 0.5 μm (10 to 20 μin.) R_a. Many plasma spray ceramic coatings can be lapped down to 0.025 to 0.05 μm (1 to 2 μin.) R_a.

Coatings for Friction and Wear Applications

The most outstanding feature of thermal spray coatings is probably their diverse applicability. There are two primary reasons for this (Ref 4). First, materials selection is almost unlimited. Second, the thermal spray process, properly controlled, imparts very little heat to the substrate (100 to 260 °C, or 200 to 500 °F), avoiding metallurgical change, distortion, and oxidation. This allows thermal spray coatings to be applied to practically any substrate (metals, plastics, composites, etc.). Also, such coatings can often be applied to the finish machined part, a practice unthinkable with welding, heat treating, or other high-temperature processes.

Thermal spray coatings provide the design engineer with many options. They offer material properties not available in wrought metals. Coatings can be applied to selected areas rather than treating the entire part. Manufacturing costs can be reduced by eliminating unnecessary processing steps. Extraordinary performance characteristics can be designed into a part to extend its useful life, creating a new, marketable product in the process.

Thermal spray coatings provide the solution to many mechanical, electrical, and corrosion resistance problems involving metal parts and assemblies. However, there are certain applications where such coatings should not be used. Before a thermal spray coating is specified, its suitability can usually be determined according to these criteria (Ref 5):

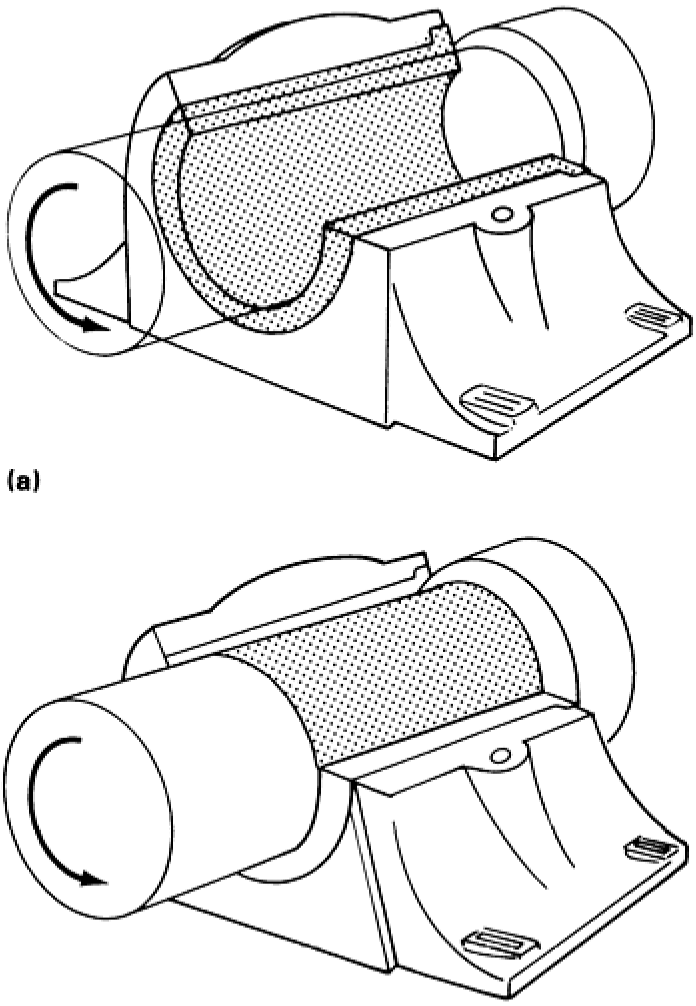
- No strength is imparted to the base material by a sprayed deposit. The component to be sprayed must, in its prepared form, be able to withstand any mechanical loading that will be experienced in service. (In a few applications, some strength can be added by a thermal spray coating; however, such uses are unusual and should be carefully tested)
- If the area on a part to be sprayed or any section of the total area will be subjected to shear loading in service, the part is not a suitable candidate for thermal spraying. Gear teeth, splines, and threads are examples
- Point loading with line contact on a sprayed metal deposit will eventually spread the deposit, causing detachment. If the deposit is on a moving component with such loading, it will fail rapidly. For example, needle and roller bearing seats, where the bearing elements are in direct contact with the sprayed deposit, may not be good thermal spray candidates
- If the base metal of a component to be treated has been nitrided, thermal spraying is not recommended unless the nitriding is first removed. Other forms of substrate hardening require special treatment, such as intensive grit blasting

Wear applications for thermal spray coatings can be categorized as adhesive wear, abrasive wear, or surface fatigue.

Adhesive Wear

Adhesive wear occurs when two surfaces slide against each other with intended motion, producing fragments from one surface that adhere to the other (Ref 6). It arises from the strong adhesive forces set up when two materials come into intimate contact. It generally occurs when lubrication is inadequate and results in metal transfer, usually called galling.

Soft bearing coatings allow the embedding of abrasive particles and also permit deformation for alignment of bearing surfaces. Adequate lubrication is required. Coatings of this type are generally low in cost, because they wear in preference to the mating surface (Fig. 6a).



(a)

(b)

Fig. 6 Cutaway views of thermal spray coated bearings showing areas that exhibit adhesive wear. (a) Soft bearing material. (b) Hard bearing material

The following performance factors apply to soft bearing coatings:

- Good lubrication must be provided or the wear rate will be excessive
- The coating must be soft enough to trap the many abrasive particles that will be carried by the lubricant
- These coatings generally have poor abrasive wear resistance
- The inherent nature of thermal spray coatings enhances their usefulness as bearing coatings. Pores act as reservoirs for lubricant; with reduced particle junctions, there is less tendency for adhesive wear

Applications and recommended materials for soft bearing coatings are listed in Table 1.

Table 1 Thermal spray coatings for friction and wear applications

Type of wear	Coating material	Coating process [(a)]	Applications
Adhesive and abrasive wear			
Adhesive wear	Soft bearing coatings:		
	Aluminum bronze	OFW, EAW, OFP, PA, HVOF	Babbitt bearings, hydraulic press sleeves, thrust bearing shoes, piston guides, compressor crosshead slippers
	Tobin bronze	OFW, EAW	
	Babbitt	OFW, EAW, OFP	
	Tin	OFW, EAW, OFP	
	Hard bearing coatings:		
	Mo/Ni-Cr-B-Si blend	PA	Bumper crankshafts for punch press, sugar cane grinding roll journals, antigalling sleeves, rudder bearings, impeller shafts, pinion gear journals, piston rings (internal combustion); fuel pump rotors
	Molybdenum	OFW, EAW, PA	
	High-carbon steel	OFW, EAW	
	Alumina/titania	OFP, PA	
	Tungsten carbide	OFP, PA, HVOF	
	Co-Mo-Cr-Si	PA, HVOF	
	Fe-Mo-C	PA	
Abrasive wear	Aluminum oxide	PA	Slush-pump piston rods, polish rod liners, and sucker rod couplings (oil industry); concrete mixer screw conveyors; grinding hammers (tobacco industry); core mandrels (dry-cell batteries); buffing and polishing fixtures; fuel-rod mandrels
	Chromium oxide	PA	
	Tungsten carbide	PA, HVOF	
	Chromium carbide	PA, HVOF	
	Ni-Cr-B-SiC/WC (fused)	OFP, HVOF	
	Ni-Cr-B-SiC (fused)	OFP, HVOF	
	Ni-Cr-B-SiC (unfused)	HVOF	
Surface fatigue wear			
Fretting: Intended motion applications	Molybdenum	OFW, PA	Servomotor shafts, lathe and grinder dead centers, cam followers, rocker arms, piston rings (internal combustion), cylinder liners
	Mo/Ni-Cr-B-SiC	PA	
	Co-Mo-Cr-Si	PA, HVOF	
Fretting: Small-amplitude oscillatory displacement applications at low temperature (<540 °C, or 1000 °F)	Aluminum bronze	OFW, EAW, PA, HVOF	Aircraft flat tracks (air-frame component); expansion joints and midspan supports (jet engine components)
	Cu-Ni-In	PA, HVOF	
	Cu-Ni	PA, HVOF	
Fretting: Small-amplitude oscillatory displacement applications at high temperature (>540 °C, or 1000 °F)	Co-Cr-Ni-W	PA, HVOF	Compressor air seals, compressor stators, fan duct segments and stiffeners (all jet engine components)
	Chromium carbide	PA, HVOF	
Erosion	Chromium carbide	PA, HVOF	Exhaust fans, hydroelectric valves, cyclone dust collectors, dump valve plugs and seats, exhaust valve seats
	Tungsten carbide	PA, HVOF	
	WC/Ni-Cr-B-SiC (fused)	OFP, HVOF	
	WC/Ni-Cr-B-SiC (unfused)	OFP, HVOF	
	Chromium oxide	PA	
Cavitation	Ni-Cr-B-SiC-Al-Mo	PA	Wear rings (hydraulic turbines), water turbine buckets, water turbine nozzles, diesel engine cylinder liners, pumps

	Ni-Al/Ni-Cr-B-SiC	PA	
	Type 316 stainless steel	PA	
	Ni-Cr-B-SiC (fused)	OFP, HVOF	
	Ni-Cr-B-SiC (unfused)	HVOF	
	Aluminum bronze	PA, HVOF	
	Cu-Ni	PA, HVOF	

(a) OFW, oxyfuel wire spray; EAW, electric arc wire spray; OFP, oxyfuel powder spray; PA, plasma arc spray; HVOF, high-velocity oxyfuel powder spray

Hard bearing coatings are highly resistant to adhesive wear. They are used where embeddability and self-alignment are not important and where lubrication is marginal (Fig. 6b).

The following performance factors apply to hard bearing coatings:

- Lubrication should be good, but is not as important as for soft bearing coatings because the high wear resistance of these materials allows them to withstand momentary unlubricated service
- Applications that require hard bearing coatings are usually characterized by high load and low speed
- Surfaces should generally be of equal hardness
- Although like coatings can be used for sliding against each other, unlike combinations are frequently used--for example, a coating running against a wrought metal. This reduces seizing and scuffing
- Wear rate generally increases with temperature

Applications for hard bearing coatings are listed in Table 1. Recommended coatings include nickel-, iron-, cobalt-, and molybdenum-base alloys, ceramics, and tungsten carbides (see Table 1).

Abrasive Wear

Abrasive wear occurs when hard foreign particles, such as metal debris, metallic oxides, and dust from the environment, are present between rubbing surfaces (Fig. 7). These particles abrade material off both surfaces. Selection of coating materials for this application should generally be based on operating temperature and surface finish requirements. The following performance factors must be considered:

- The coating must be hard. In particular, surface hardness should exceed the hardness of the abrasive grains present
- The most common abrasive is silica (sand), with a hardness of approximately 820 HK. (For comparison, tungsten carbide/cobalt composite is 1400 to 1800 HK; Al_2O_3 is approximately 2100 HK)
- Information about the abrasive--how often it is replenished, whether it is sharp and brittle, how it breaks down--is important in selecting the coating and estimating its performance
- If the system is closed, debris created by the wear process will also contribute to the wear rate and thus must also be considered
- The coating must exhibit oxidation resistance at the service temperature

Applications and recommended materials for coatings resistant to abrasive grains at low and high temperatures are listed in Table 1.

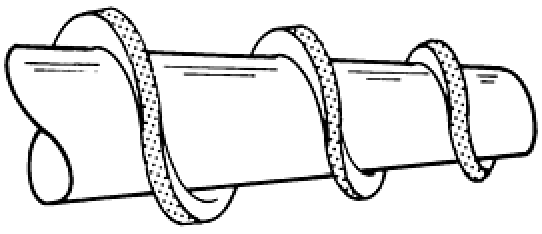


Fig. 7 Abrasive wear on the flight surfaces of a helical screw used in spiral conveyor applications

Surface Fatigue Wear

Repeated loading and unloading cause cyclic stress on a surface, eventually resulting in the formation of surface or subsurface cracks. The surface ultimately fractures and large fragments are lost, leaving pits. This phenomenon can occur only in systems where abrasive and adhesive wear are not present--for example, in systems with high surface contact loads. An area of surface must be stressed repeatedly, without constant removal of particles, to fail in fatigue. Fretting, erosion, and cavitation are typical examples of this type of wear.

Fretting. Some fretting-resistant coatings resist wear caused by repeated sliding, rolling, or impacting over a track (Fig. 8a). The repeated loading and unloading cause cyclic stresses inducing surface or subsurface cracks. Other coatings resist wear caused when contacting surfaces undergo oscillatory displacement of small amplitude (Fig. 8b). This type of wear is difficult to anticipate, because no intended motion is designed into the system. Vibration is a common cause of fretting. The following performance factors apply to coatings for fretting resistance:

- The coating must be resistant to oxidation at the service temperature. If an oxide forms, it must be tough and tenacious; a loosely adherent oxide will cause severe abrasive wear
- A surface that is free of stress, particularly tensile stress, is desirable. High-shrink coatings tend to have high surface stress and do not perform as well as low-stress coatings
- Brittle coatings fail rapidly. Tough coatings tend to perform better
- Coatings with hard particles distributed in a soft matrix are generally the most durable

Applications and recommended materials for fretting-resistant coatings are listed in Table 1.

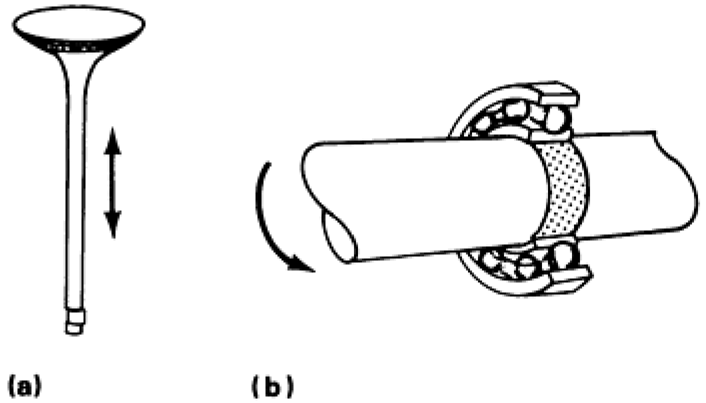


Fig. 8 Two types of fretting wear. (a) Stem and seat wear caused by the intended cyclic up-and-down motion of an engine valve. (b) Wear caused by the unplanned but unavoidable oscillatory motion of a press-fitted shaft on the inner ring of a bearing

Erosion is caused when a gas or a liquid that ordinarily carries entrained particles impinges on a surface with velocity (Fig. 9). When the angle of impingement is small, the wear-producing mechanism is closely analogous to abrasion. When the angle of impingement is normal to the surface, material is displaced by plastic flow or is dislodged by brittle failure. The following performance factors apply to erosion-resistant coatings:

- If the angle of particle impact is less than 45°, the coating selected should be harder and more abrasion resistant
- If the angle of particle impact is greater than 45°, the coating should be softer and tougher
- At high service temperatures, coatings should have high hot hardness and oxidation resistance at temperatures and environments ranging from 540 to 815 °C(1000 to 1500 °F)
- When the carrier is liquid, the corrosion resistance of the coating must be considered

Applications and recommended materials for coatings used to resist particle erosion are listed in Table 1.

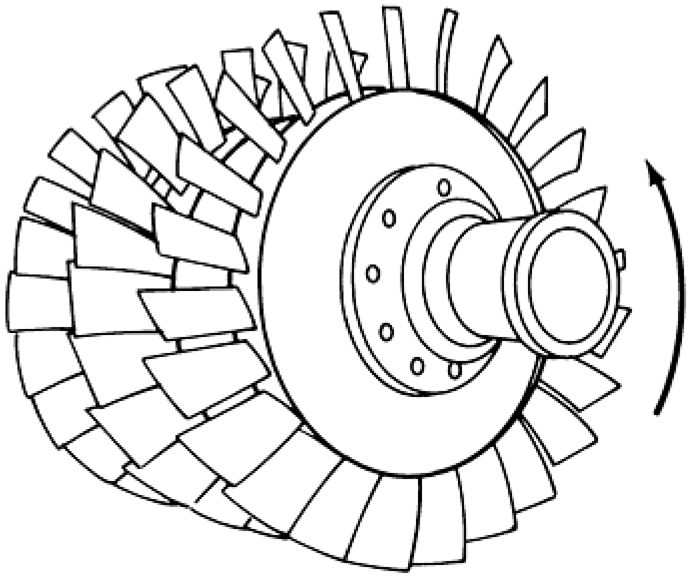


Fig. 9 Typical erosive wear of a fan blade assembly generated by the high-velocity impingement of a gas or liquid on the blade surface

Cavitation is caused by mechanical shock that is induced by bubble collapse in liquid flow (Fig. 10). Materials that resist fretting-type surface fatigue are resistant to cavitation. The most effective coating properties are toughness, high wear resistance, and corrosion resistance. The following performance factors apply to cavitation-resistant coatings:

- Relative motion between a liquid and metal surface, including bubble generation and bubble collapse, must exist for cavitation to occur
- Liquids will penetrate sprayed coatings unless fused; therefore, all coatings should be sealed
- Selection of a coating must be influenced by its resistance to the liquid used in a particular application
- Hardness is an important factor, but coatings must also be tough. Brittle coatings fail quickly
- Coatings that work harden are especially resistant to the repeated pounding of cavitation

Applications and recommended materials for coatings resistant to cavitation are listed in Table 1.

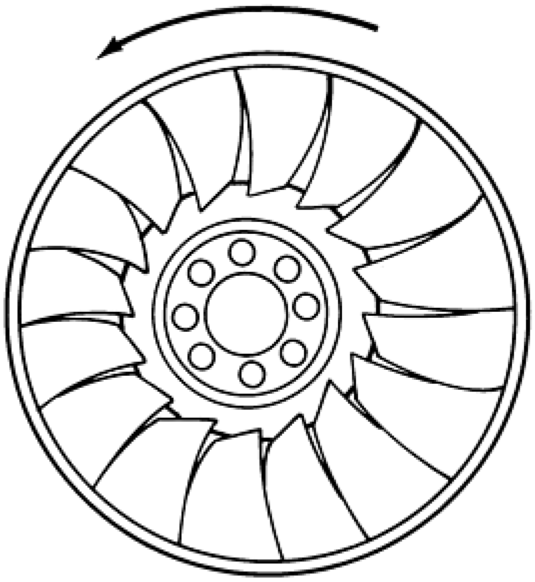


Fig. 10 **Typical surface fatigue wear produced by cavitation in a pump impeller component**

References

1. E.R. Novinski, Application of Thermal Spray Processes to Conserve Critical Materials, *Workshop on Conservation and Substitution Technology for Critical Materials, Conference Proceedings,* Vanderbilt University, 16 June 1981
2. H. Herman, Plasma Sprayed Coatings, *Sci. Am.,* Sept 1988
3. H.S. Ingham and A.P. Shepard, *Flame Spray Handbook,* Vol III, Perkin-Elmer Corp., Metco Division, 1965
4. G.L. Kutner, Thermal Spray by Design, *Adv. Mater. Proc. Met. Prog.,* Oct 1988, p 67
5. M.L. Thorpe, Thermal Spraying Becomes a Design Tool, *Mach. Des.,* 24 Nov 1983
6. F.N. Longo, *Handbook of Coating Recommendations,* Perkin-Elmer Corp., Metco Division, 1972

Electroplated Coatings

Rolf Weil and Keith Sheppard, Stevens Institute of Technology

Introduction

IT IS OFTEN NECESSARY to coat a material that is subject to friction and, possibly, wear. The coating can be tailored to the tribological demands of the environment and can provide a wider choice in selecting a base material to meet special requirements, such as strength or low cost.

Like other of surfaces used in tribological applications, electroplated coatings have two primary categories. Hard coatings are normally used to resist many forms of wear, such as those that involve abrasive, adhesive, and erosion process. Some degree of toughness in these coatings is often desirable, in order to resist cracking. Soft coatings are sometimes used on bearing surfaces to provide low shear strength. They are typically used at ambient temperatures and low loads.

An additional consideration in coating selection is a requirement to provide corrosion protection. Corrosive wear places particularly severe demands on the protective abilities of such a coating, because the thin surface film that is primarily responsible for limiting the corrosion kinetics is continually being worn away.

Coatings that are used to control friction and wear can be electrochemically deposited either with or without an externally applied current. Deposition without an external current is called electroless plating. For many wear applications, electrochemical deposition is the most rapid and economical means to apply coatings ranging from 10 to 500 μm (0.4 to 20 mils) in thickness (Ref 1).

Adhesion to the substrate is a very important requirement of all coatings. Because adhesion primarily depends on the cleanliness of the substrate surface to be coated, proper pretreatment of the substrate is usually necessary.

Deposition Fundamentals

Electroplating is the coating of an electrically conducting surface by application of an electrical potential in a suitable solution that contains the ions of the metals to be deposited. The electrode to be coated is the cathode. The counter-electrode, the anode, can be of the soluble type, so that it supplies metal ions to the solution. Alternatively, the anode can be insoluble, in which case the ions of the metals to be deposited must be continuously or periodically added to the plating solution to compensate for the depletion.

The deposition rate depends primarily on the current density. If all supplied electrons reduce the metal ions, then the deposition rate can be readily calculated from Faraday's law, which states that 96,500 coulombs (1 Faraday) deposit 1 gram (0.035 oz) equivalent weight (atomic weight divided by valence). However, hydrogen evolution or other secondary reactions may use some of the current supplied. Thus, only a fraction of the supplied electrons reduces the metal ions. In this case, the plating efficiency is less than 100%. (The plating efficiency is the ratio of the metal yield to that calculated from Faraday's law, which assumes that there are no other reactions.)

The microstructures of electrodeposits, which to a large extent determine their properties, depend on a number of factors. These factors include the microstructure of the surface to be coated, the plating conditions (that is, the current density), the temperature and composition of the plating solution, as well as the degree and type of agitation. The composition and the pH of the plating solution in the vicinity of the cathode (which generally differs from that of the bulk) can have a large effect on the structure and properties of the deposit. In the absence of significant surface inhibition, it is possible for the deposit to reproduce the structure of the substrate surface. However, even minute quantities of certain substances can greatly affect the structure. These substances may be intentionally added to the plating solution or be present as impurities from the water, chemicals, or secondary reactions. Foreign substances can alter the grain orientation by being preferentially adsorbed on certain crystal planes. Then, grains with other planes exposed on their surfaces can grow preferentially until they essentially compose the entire deposit, thereby producing a fiber texture, that is, most of the grains have same crystallographic direction normal to the surface.

The adsorption of some foreign material can greatly reduce the grain size. These materials impede grain growth, thereby requiring almost continuous renucleation. Some addition agents are present in plating solutions to level the deposit surface, that is, to make it smoother than the substrate. They do so by being preferentially adsorbed on asperities. By blocking growth on the asperities, the recesses receive most of the depositing metal, thereby producing a smoother surface.

High internal stresses, particularly of the tensile type, can adversely affect the wear properties of electrodeposits. They can arise from the lattice misfit between the substrate and an epitaxial deposit, from the coalescence of crystallites, or from the diffusion of hydrogen out of the surface layers. These stresses can be large enough to cause cracking, as is the case in some chromium deposits. Most electrodeposits are also characterized by a high dislocation density, which results in strength, hardness, and ductility values that are similar to those of the same metal in the cold-worked condition.

Pulse plating has been used to enhance the tribological properties of electrodeposits. Generally, a square-wave current pulse is employed. The current is on for a certain time period and off for another. The ratio of the time to the sum of the on and off times is called the duty cycle. By varying the pulse frequency, the duty cycle, and the current density, the structure and properties of electrodeposits can be beneficially changed.

Electroless plating is an autocatalytic reaction in which a reducing agent supplies an internal current. The advantages of electroless plating are that nonmetallic substrates can be coated and the deposit thickness is uniform. The autocatalytic

reaction has to be activated on some surfaces, particularly those that are nonmetallic. Thus, the areas to be plated can be controlled by specific activation. The disadvantages are a slow deposition rate and a higher cost. It is also necessary to maintain the composition of the plating solution by continuous or frequent additions of the depleted chemicals and removal of the impurities.

Wear-resistant composite can be produced by the codeposition of fine particulate matter. Hard particles, such as diamond, silicon carbide, and aluminum oxide, are kept in suspension by agitation and become occluded in the deposit. Solid lubricants, such as polytetrafluoroethylene (PTFE) particles, have also been codeposited. As much as 30 vol% of the particles can be attained in both electrodes and electroplated coatings (Ref 2).

In some applications, only limited areas of a component are subjected to wear. If the deposition of precious metals is involved, then it may be economically desirable to coat these areas selectively. Such limited coverage can be achieved by jet plating, laser-enhanced and laser-jet plating, and by physically masking off areas not to be plated, as in photolithography (Ref 3).

In jet plating, a fine jet of the plating solution is directed onto the areas to be coated. Because the current is constrained within the jet, only the areas where it is applied become coated. The jet also provides rapid ion transport to the depositing surface and therefore permits to the deposition rates. In laser-enhanced plating, the heating effect of the laser enhances mass transport locally. It can also influence the deposition kinetics such that metal is not plated outside of the heated area. Laser-jet plating, which is a combination of jet and laser-enhanced plating, can provide improved deposit characteristics. The jet then acts as a waveguide for the laser.

The principal electrochemically deposited materials used in tribological applications are chromium, nickel, and both precious and soft metals. The characteristics of each type of deposit are described in this article. References 4, 5, 6, 7, 8, 9, 10, 11 provide additional information relevant to electrochemically deposited metals and alloys used in tribological applications.

Chromium

Hard chromium coatings are widely used because of their low coefficient of friction and good wear properties. They are deposited at higher temperatures and current densities than decorative chromium. The plating solution for hard chromium has a lower ratio of chromic oxide to sulfuric acid (the main constituents of the solution) than that used for decorative coatings. The thickness of hard chromium deposits varies from about 0.1 to 100 μm (0.004 to 4 mils), whereas decorative thicknesses usually range from about 0.1 to 0.2 μm (4 to 8 μin.). Very strict antipollution to regulations govern the discharge of solutions containing hexavalent chromium ions, which are used for hard-chrome plating. Therefore, plating solutions that contain mostly trivalent chromium ions are of interest. These solutions generally contain formic acid or one of its salts. Carbon will deposit in these solutions. The deposits can therefore be heat treated to precipitate a chromium carbide.

The hardness of hard chromium varies from about 900 to 1100 on the Knoop and Vickers hardness scales. These values are considerably higher than the hardness of bulk chromium. Deposits from trivalent solutions are softer than those that are plated from hexavalent chromium solutions. However, after heat treating at about 700 °C (1290 °F), a hardness comparable to that of hard chromium can be achieved (Ref 12).

Chromium deposits are characterized by high internal tensile stresses that can reach 1000 MPa (145 ksi). These stresses can reduce the fatigue properties of coated components. Hydrogen is also codeposited with chromium and can diffuse into components, causing hydrogen embrittlement. Heat treatments are typically required to relieve the stresses and hydrogen effects, but can reduce the hardness.

The coefficients of friction of hard chromium against hard materials are generally the lowest of any electrochemically deposited coatings. The actual values vary considerably, depending on the test method, the mating surfaces of the materials, and the degree of lubrication. Some values of static and sliding coefficients of friction are listed in Table 1. The static coefficient is calculated from the force to initiate movement of one component of a couple against the other. The force to maintain movement enters into the calculation of the sliding coefficient. It is important to note that only coefficients of friction obtained under the same conditions can be compared. The values should not be considered as absolute.

Table 1 Coefficients of friction

Couple	Static coefficient	Sliding coefficient
Chromium-plated steel versus itself	0.14	0.12
Chromium-plated steel versus steel	0.15	0.13
Steel versus steel	0.30	0.20

Source: Ref 13

The wear rates of hard chromium can vary greatly, depending, again, on the type of test. The rates also vary with the mating material and whether adhesive or abrasive wear predominates. Some dry abrasive wear data from a Taber abrasion test for three chromium deposits labeled CrA, CrB, and CrC are shown in Fig. 1. The Taber test measures, for a certain number of cycles, the weight loss that results from abrasion with resilient, abrasive wheels at a load of 9.8 N (11 kgf). Figure 1 shows that deposit CrC had less wear than deposit CrA.

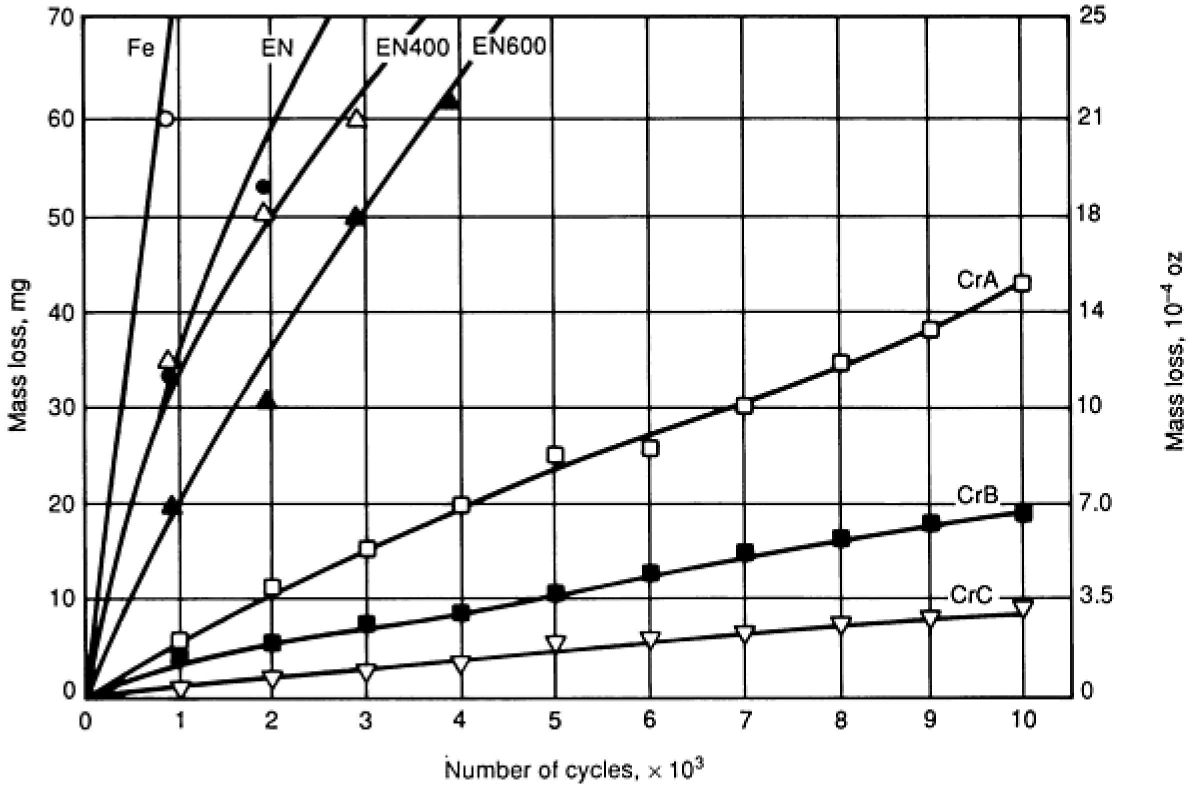


Fig. 1 Effect of number of cycles on mass loss in the Taber abrasion test for uncoated steel substrate (Fe), three chromium deposits (CrA, CrB, CrC), and three electroless nickel deposits: as-plated nickel (EN), heat treated at 400 °C (750 °F) (EN400), and heat treated at 600 °C (1110 °F) (EN600). Source: Ref 1

Figure 2 represents data obtained in a Falex wear test for the same three chromium deposits. In this test, a pin is rotated between two V-shaped blocks. Deposit CrA showed less wear than deposit CrC, illustrating the effect of a lubrication test method on the results. Figure 1 also shows that the chromium coating improved the abrasion resistance of the steel substrate. The hardest chromium deposits do not necessarily exhibit the least wear. The low friction coefficients and good wear properties of chromium have been attributed to a self-healing Cr_2O_3 film that forms on the surface. In general, hard chromium has a lower wear rate than either electroplated or electroless nickel, which are the two competing materials. This effect is also illustrated in Fig. 1 and 2.

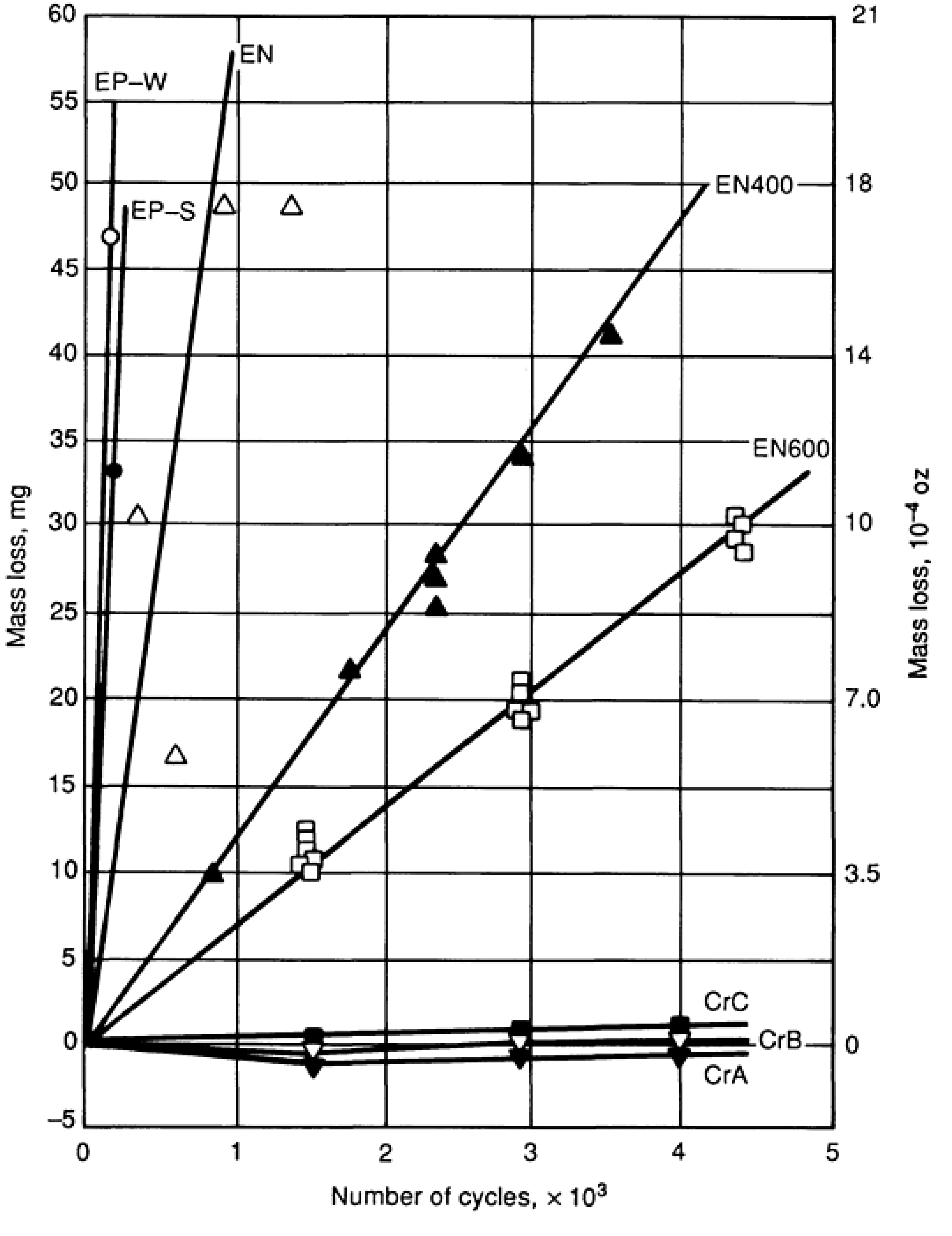


Fig. 2 Effect of number of cycles on mass loss of plated pin versus steel blocks in a Falex test for same three chromium deposits (CrA, CrB, CrC) and same three electrodes nickel deposits (EN, EN400, EN600) shown in Fig. 1. Effects on two electroplated nickles from a sulfamate solution (EP-S) and a Watts solution (EP-W) are shown. Source: Ref 1

Under corrosive wear conditions, chromium coatings do not protect substrates if they crack in response to high internal stresses. Pulse plating can reduce the internal stresses and the resulting cracking of the deposit. The effects of pulse plating and deposition temperature on the wear rate of chromium deposits are shown in Fig. 3. The improve wear resistance at higher pulse frequencies and temperatures corresponds to increases in hardness. The wear resistance in some applications can also be improved by the inclusion in the deposit of hard particles or those of a solid lubricant (Ref 15).

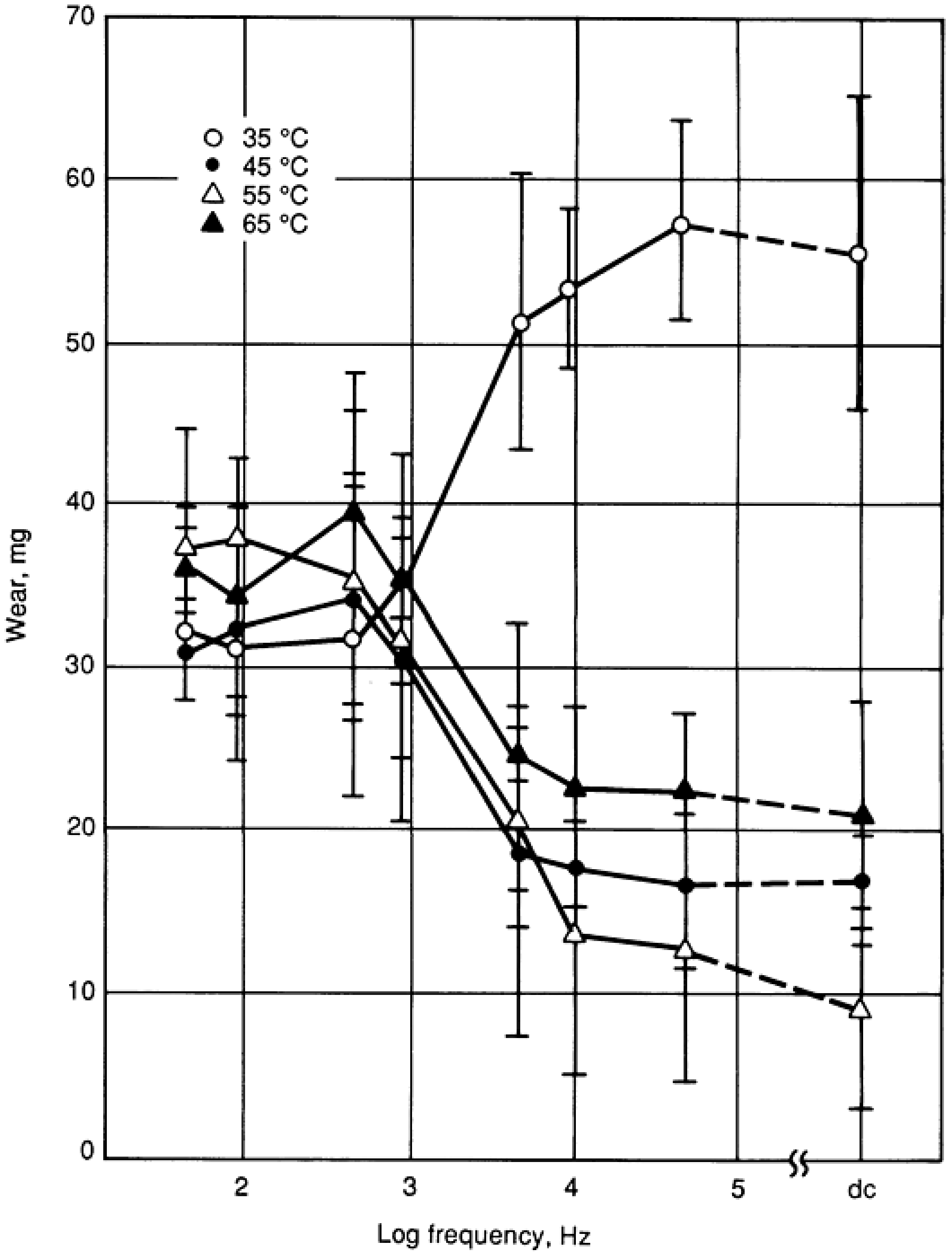


Fig. 3 Effect of pulse frequency and solution temperature on mass loss of chromium deposits in a Taber abrasion test. dc, direct current. Source: Ref 14

Chromium is widely used for wear resistance in automotive and aircraft components, such as pistons and shock absorbers (Ref 13). Other applications include coatings on drills, taps, dies, extrusion screws, and rolls. The wear resistance of gun barrels can also be improved by chrome plating. Salvaging of worn parts by chromium electrodeposition is an important industrial application.

Most of these applications require relatively thick chromium deposits. Thick deposits have a nodular surface structure, which is shown in Fig. 4. Cracks that are due to high internal stresses are also visible. The nodular structure is too rough for some applications and therefore requires a mechanical finishing operation. Machining tools are usually plated with a thin chromium deposit, which is smooth and does not need to be finished (Ref 16).

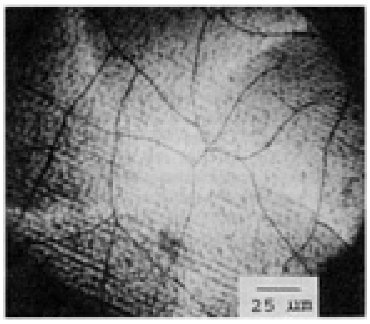


Fig. 4 Optical micrograph of electrodeposited chromium showing nodular structure and cracks

Electroplated Nickel

The most widely used solution for plating nickel for wear applications is the Watts solution. Its main components are nickel sulfate, nickel chloride, and boric acid. Organic addition agents in the plating solution can increase the hardness and wear resistance mainly by decreasing the grain size. Nickel usually is deposited with a tensile internal stress. Some sulfur compounds can cause the stress to become compressive, but also make the deposit more brittle, especially under elevated-temperature conditions. The nickel sulfamate plating solution produces low-stress deposits. It is possible to codeposit such metals as tungsten and molybdenum with nickel, even though they cannot be plated alone in aqueous solutions. Inclusion of hard particles or those of solid lubricants can also improve the wear or friction properties of electroplated nickel.

Improved wear resistance resulting from the incorporation of SiC particles is shown in Fig. 5. In the test on which the data of Fig. 5 are based, a plated block was pressed against a lubricated steel ring. Further improvement that was done to phosphide precipitates in heat-treated Ni-P-SiC coatings is also seen in this figure.

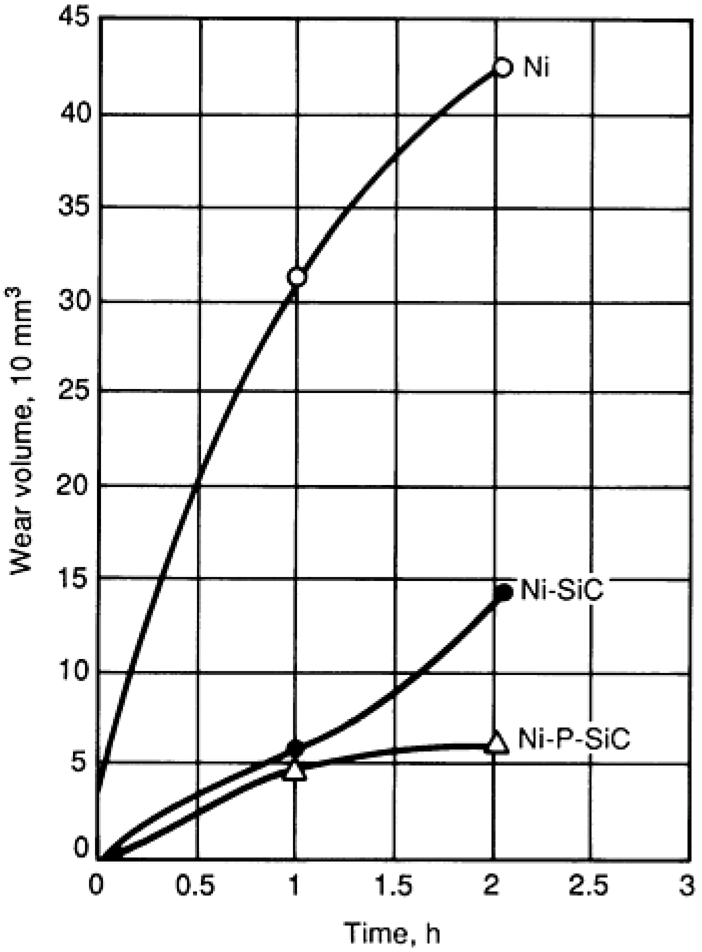


Fig. 5 Effects of codeposited SiC particles and phosphide precipitates on wear of electrodeposited nickel heat treated for 1 h at 400 °C (750 °F). Source: Ref 17

The hardness of nickel deposits can vary from about 150 to 500 on the Vickers scale. The hardness depends on the plating conditions, that is, current density, solution pH and temperature, and composition. Pulse plating can increase the hardness. The coefficient of friction and wear rates of electroplated nickel are generally greater than those of chromium. The lower wear rates of electroless, compared with electroplated, nickel can be seen in Fig. 1 and 2. The widest use of electroplated nickel for wear applications is as an undercoat for chromium. If thick deposits are needed, for example, in building up heavily worn parts, it is usually not practical to do so using only chrome plating, because of its low current efficiency ad high internal stress. In such cases, most of the deposit thickness is composed of nickel, with chromium constituting only a thin outer layer.

Under corrosive conditions, chromium may not protect the substrate, because of its cracks, but a nickel undercoat will. There are some applications of electrodeposited Ni-W, Ni-Mo, and Ni-Cr alloys that are hard and wear resistant. These alloys also offer good corrosion protection.

Electroless Nickel

The tribological and associated properties of electroless nickel were extensively discussed by Weil and Parker (Ref 18). The reducing agents for electroless nickel are sodium hypophosphite, sodium borohydride, or organic aminoboranes. When sodium hypophosphite is the reducing agent, the deposit generally contains between 3 and 11 wt% phosphorus. The boron contents of electroless nickel range from 0.2 to 4 wt% and from 4 to 7 wt% when the reducing agents are an aminoborane and sodium borohydride, respectively.

As-deposited electroless nickel is a supersaturated solid solution because the equilibrium solid solubility of boron or phosphorus in nickel is essentially zero. However, the second phase cannot form, because the time interval between the

deposition of successive layers is too short for the necessary diffusion. The structure of the deposit changes from microcrystalline to amorphous, with increasing alloy content. The second phase, which is Ni_3P Ni_3B, can form on annealing and result in precipitation hardening.

Electroless nickel can be heat treated to hardnesses comparable to those of electrodeposited chromium. The maximum hardness can be attained in 1 h at about 400 °C (750 °F) or 10 h at 260 °C (500 °F). The hardness of as-plated Ni-P alloys varies from 500 to 650 on the Vickers scale. As-plated Ni-B deposits are generally harder than the Ni-P ones. The ability of electroless nickel deposits to maintain their hardness under elevated-temperature service conditions increases with increasing phosphorus or boron content, but decreases rapidly above 385 °C (725 °F). Nickel-boron coatings tend to better withstand wear at elevated temperatures and are therefore more widely used under these conditions.

Coefficients of friction of electroless nickel in the as-deposited condition (EN) and heat treated at 400 °C (750 °F) (EN400) and at 600 °C (1110 °F) (EN600) are listed in Table 2. They are compared to the three chromium alloys depicted in Fig. 1 and 2. The counter surfaces were diamond and plain carbon steel. The coefficients friction of the electroless nickels are higher than those of the chromium deposits. It should be noted that the values for the chromium deposits against steel are considerably higher than those listed in Table 1, again highlighting the fact that only data obtained under identical conditions can be compared.

Table 2 Coefficients of friction of chromium versus electroless nickel

Coating	Coefficients of friction	
	Counterface diamond	**Counterface plain C steel (080M40)**
CrA	0.040	0.88
CrB	0.035	0.82
CrC	0.030	0.81
EN	0.180	0.96
EN400	0.300	0.95
EN600	0.060	0.90

Source: Ref 1

Electroless nickel is most frequently used in wear applications in the precipitation-hardened condition. The improved abrasion and wear resistance after heat treating is evident from Fig. 1 and 2. Parker (Ref 19) found that the heat-treated electroless nickel exhibited smaller weight losses for comparable numbers of cycles than are shown in Fig. 1. Although the greatest hardness was obtained by heat treatment at 400 °C (750 °F), the electroless nickel that was heat treated at 600 °C (1110 °F) showed the least wear. However, the corrosion resistance of as-plated Ni-P coatings, which is one of the main reasons for their use, deteriorates upon heat treating to the maximum hardness, because of cracking. Therefore, under corrosive wear applications, it may be prudent for forgo the higher hardness.

The widest use of electroless nickel for both corrosion and wear resistance is in valves that control the flow of either liquids or gases. Other uses (Ref 20) of electroless nickel in wear applications include aluminum piston heads, aircraft engine shafts, components of gas turbines and engine mounts in the aircraft industry, and such automotive parts as differential pinion ball shafts, fuel injectors, ball studs, disk brake pistons, transmission thrust washers, knuckle pins, and hose couplings.

In mining applications and associated material-handling equipment, where abrasive wear conditions prevail, electroless nickel has been substituted for hard chromium. A very important application of electroless nickel is the salvage of worn surfaces, especially because it is possible to coat only specific areas. Thus, little or no subsequent machining or grinding is required. The inclusion in electroless-nickel deposits of hard or solid-lubricant particles can also improve their tribological properties. The effect of these particles on the wear of electroless nickel is similar to that shown in Fig. 5.

Precious Metal Deposits

Separable electrical connectors are not only subjected to wear, but must maintain low contact resistance. Electrodeposited gold from a cyanide plating solution is most widely used, because it does not form on oxide surface film of high electrical resistivity. Soft gold deposits wear poorly. The addition of cobalt ions to the plating solution is the most frequent means

used to increase the hardness of gold. However, cobalt additions reduce the ductility of the deposits. These effects appear to be due primarily to grain-size reduction. The addition of cobalt is accompanied by the codeposition of fine polymer particles that are cobalt-cyanide compounds.

Gold plate from solutions that contain nickel ions is also used for separable connectors. An additive-free, gold-plating solution based on phosphoric acid results in deposits that have properties comparable to those of the cobalt-hardened deposits (Ref 21). To prevent interdiffusion of gold and the substrate, which is usually copper, a layer of nickel is often deposited as a barrier.

Palladium, Pd-Ag, and Pd-Ni have been substituted for gold (Ref 22), primarily in an effort to reduce costs. Polymerization of organic air pollutants can occur on the catalytically active palladium surface under tribological conditions and result in a film of high electrical contact resistance. Alloying reduces polymer formation. Other measures can also be taken, including coating one or both mating surfaces with a thin gold layer (250 μm, or 10 mils) or using higher contact forces. Fretting corrosion, which results from the small-amplitude movements of the mating surfaces that are often due to vibrations or different thermal expansion rates, can cause a thin gold deposit to wear through and the polymer film to form.

The Knoop hardness is about 200 for cobalt-hardened gold and about 180 for gold from the additive-free phosphate solution. On the same scale, 25-karat gold has a hardness of only about 50. Table 3 compares the characteristics of various precious metal contacts (Ref 23). A low ductility, that is, less than 1% elongation in a tension test, is desirable for gold deposits to reduce prow formation, which is the build-up of a lump of work-hardened metal. Antler (Ref 24) has reviewed the tribology of electronic connectors.

Table 3 Plated contact metallurgical properties

Properties	Gold		Palladium	Palladium-nickel
	Soft	**Hard**		
Resistivity, μΩ · cm	2.1-2.3	3.8-4.2	9.9-10.7	n/a
Contact resistance, mΩ	0.7	1.4	3.5	2.0[a]
Hardness, kgf/mm^2	60-100	130-200	200-400	500-650
Density, g/cm^3[b]	19.3	17	12	11
Corrosion resistance	E	E	VG	E[a]
Solderability	E	P	G	VG[a]
Wear resistance	P	G	VG	E[a]
Cost[b] (equal thickness)	. . .	. . .	30% of Au	20% of Au

E, excellent; VG, very good, G, good; P poor.

(a) With gold flash.

(b) Approximate values.

Soft Metals

Electrodeposited soft metals such as tin, lead-tin, and silver alloys are used as bearing coatings. Tin and silver deposits prevent galling, especially during start-up. The codeposition of 3.4 to 3.6% Pb with silver provides excellent antiseizure coatings for bearings (Ref 25). Silver alloys are used for elevated-temperature solid lubrication. Separable electrical contacts have also been coated with electrodeposited Pb-Sn. In this case, friction results in the removal of a high-resistivity surface film. An alloy of lead with from 4 to 10% tin, and sometimes containing copper, is used for sleeve-bearing overlays (Ref 26). Lead alloyed with indium is an excellent bearing material that also possesses good corrosion resistance. However, it is more expensive than alloys of lead and tin. Electrodeposits for bearings are reviewed in greater detail by Eastham and Crooks (Ref 27).

Magnetic Materials

Electroless and electrodeposited magnetic alloys, such as Permalloy (Ni-Fe) and various cobalt alloys (Co-P and Co-Ni-P), are used in advanced magnetic storage technologies, such as hard memory disks for computer systems. Electroless nickel that has a high enough phosphorus content (greater than 7 wt%) to be amorphous and therefore nonmagnetic is commonly used as an underlayer for magnetic coatings. These storage systems represent a demanding tribological environment, because of the need to maintain minimal separation between the read/write heads and the storage media. This close proximity is necessary to maximize information storage density. It has been suggested that the resources lost to friction and wear at the head-media interfaces in magnetic storage amount to about 1% of the U.S. gross national product (Ref 28). Even if this figure were an overstatement, it does indicate the scale of the problem.

Because it is necessary to limit the wear of the head and disk materials, the very thin magnetic layer on storage disks is normally coated with a thin wear-resistant and/or lubricating layer. The coating also affords corrosion resistance. Electroplated rhodium, which has properties similar to hard chromium, has been used for this purpose. However, it is now more typical to employ such coatings as silicates or the so-called diamond-like hydrogenated carbon (Ref 3). In addition to the coating, a lubricating layer may also be necessary.

Moisture and ionic contaminants, such as salts, can lead to corrosion. Porosity in the protective coating can allow corrosion of the magnetic layer and undercoat. The primary strategy is to minimize corrosion by producing pore-free protective layers. However, corrosion resistance can be improved by incorporating elements such as chromium into the magnetic layer, as long as potentially detrimental effects on the magnetic properties can be ameliorated. In practice, it is possible to codeposit chromium by vapor phase methods. Presently, the codeposition of chromium in cobalt magnetic films represents a challenge to electrodeposition technology, along with the provision of a self-lubricating surface (Ref 3). It is also apparent that the need to maintain submicron spacing between the magnetic layer and the heads in advanced storage systems requires much better control of the electrochemical plating processes than is generally exercised.

References

1. D.T. Gawne and U. Ma, Friction and Wear of Chromium and Nickel Coatings, *Wear,* Vol 129, 1989, p 123
2. N. Feldstein, Composite Electroless Plating, *Electroless Plating: Fundamentals and Applications,* G.O. Mallory and J.B. Haydu, Ed., American Electroplaters and Surface Finishers Society, 1990, p 269-288
3. L.T. Romankiw and T.A. Palumbo, Electrodeposition in the Electronic Industry, *Electrodeposition Technology, Theory and Practice,* L.T. Romankiw and D.R. Turner, Ed., *Proceedings,* Vol 87-17, The Electrochemical Society, 1987, p 13-41
4. *Metals Handbook,* 9th ed., *Surface Cleaning, Finishing and Coating,* Vol 5, American Society for Metals, 1982
5. F.A. Lowenheim, Ed., *Modern Electroplating,* 3rd ed., John Wiley and Sons, 1974
6. L. Durney, Ed., *Electroplating Engineering Handbook,* 4th ed., Van Nostrand Reinhold, 1984
7. J.K. Dennis and T.E. Such, *Nickel and Chromium Plating,* Butterworth, 1972
8. G.O. Mallory and J.B. Haydu, Ed., *Electroless Plating: Fundamentals and Applications,* American Electroplaters and Surface Finishers Society, 1990
9. J.Cl. Puippe and F. Leaman, Ed., *Theory and Practice of Pulse Plating,* American Electroplaters and Surface Finishers Society, 1986
10. W.H. Safranek, Ed., *The Properties of Electrodeposited Metals and Alloys: A Handbook,* 2nd ed., American Electroplaters and Surface Finishers Society, 1986
11. Electrodeposited Coatings Database, American Electroplaters and Surface Finishers Society
12. M. Takaya, M. Matsunaga, and T. Otaka, Hardness of Deposits from Trivalent Chromium Sulfate/Potassium Formate Baths, *Plat. Surf. Finish.,* Vol 74 (No. 6), 1987, p 90
13. G. Dubpernell, Chromium, Chapter 5, *Modern Electroplating,* 3rd ed., John Wiley and Sons, 1974, p 87-151
14. J.H. Chai, D.Y. Chang, and S.C. Kwon, The Properties of Chromium Electroplated with Pulse Current, *Plat. Surf. Finish.,* Vol 76 (No. 6), 1989, p 80
15. J. Zahavi and H. Kerbel, Properties of Electrodeposited Composite Coatings, *Plat. Surf. Finish.,* Vol 69

(No. 1), 1987, p 76
16. K.G. Budinski, Wear Characteristic of Industrial Plating, *Selection and Use of Wear Tests for Coatings,* STP 769, ASTM, 1982, p 118-131
17. X. Changgeng, D. Zonggeng, and Z. Lijun, The Properties of Electrodeposited Ni-P-SiC Composite Coatings, *Plat. Surf. Finish.,* Vol 75 (No. 10), 1988, p 54
18. R. Weil and K. Parker, Chapter 4, Properties of Electroless Nickel Plating, *Electroless Plating: Fundamentals and Applications,* American Electroplaters and Surface Finishers Society, 1990, p 111-138
19. K. Parker, Hardness and Wear Resistance Tests of Electroless Nickel Deposits, *Plat. Surf. Finish.,* Vol 61 (No. 9), 1974, p 834
20. J. Colaruotolo and D. Tramontana, Chapter 8, Engineering Applications of Electroless Nickel, *Electroless Plating: Fundamentals and Applications,* American Electroplaters and Surface Finishers Society, 1990, p 207-228
21. F.B. Kock, Y. Okinaka, C. Wolowodiuk, and D.R. Blessington, Additive-Free Hard Gold Plating for Electronic Applications, *Plat. Surf. Finish.,* Vol 67 (No. 6), 1980, p 50
22. G.L. Ide and J.B. Vanhumbeeck, Palladium and Palladium Alloy Electroplating for Contact Applications, *Electrodeposition Technology, Theory and Practice,* L.T. Romankiw and D.R. Turner, Ed., *Proceedings,* Vol 87-17, The Electrochemical Society, 1987, p 179-190
23. A.D. Knight and S. Levine, High Performance Signal Connectors and Contact Interface Metallurgy, *Electrochemical Technology in Electronics,* L.T. Romankiw and T. Osaka, Ed., *Proceedings,* Vol 88-23, The Electrochemical Society, 1988, p 161-170
24. M. Antler, The Tribology of Contact Finishes: Mechanisms of Friction and Wear, *Plat. Surf. Finish.,* Vol 75 (No. 10) 1988, p 46
25. B.M. Luce and D.G. Foulke, Silver, Chapter 14, *Modern Electroplating,* 3rd ed., John Wiley and Sons, 1974, p 358-376
26. H.J. Wiesner, Lead, Chapter 11, *Modern Electroplating,* 3rd ed., John Wiley and Sons, 1974, p 266-286
27. D.R. Eastham and C.S. Crooks, Plating for Bearing Applications, *Trans. Inst. Metal Finish.,* Vol 60, 1982, p 9
28. B. Bhushan, *Tribology and Mechanics of Magnetic Storage Devices,* Springer-Verlag, 1990

PVD and CVD Coatings

R.F. Bunshah, University of California at Los Angeles

Introduction

VAPOR DEPOSITION PROCESSES can principally be divided into two types (Ref 1):

- *Physical vapor deposition* (PVD) processes, which require creation of material vapors (via evaporation, sputtering, or laser ablation) and their subsequent condensation onto a substrate to form the film
- *Chemical vapor deposition* (CVD) processes, which are generally defined as the deposition of a solid material from the vapor phase onto a (usually) heated substrate as a result of numerous chemical reactions

PVD and CVD processes can be classified as shown in Table 1.

Table 1 Classification of physical vapor deposition (PVD) and chemical vapor deposition (CVD) processes on the basis of material deposited on substrate

Process	Metals		Compounds	
	Basic	**Hybrid**	**Basic**	**Hybrid**
Physical vapor deposition				
Conventional	Evaporation deposition	Ion plating	. . .	. . .
	Sputter deposition	. . .	. . .	. . .
Plasma-assisted (PAPVD)	. . .	. . .	Activated reactive evaporation (ARE)	Reactive ion plating (RIP)
	. . .	. . .	Reactive sputtering (RS)	. . .
Chemical vapor deposition				
Conventional	Thermal	. . .	. . .	. . .
Plasma-assisted (PACVD)	. . .	. . .	rf excitation	. . .
	. . .	. . .	Microwave excitation	. . .

Table 2 shows selected parameters of basic PVD and CVD processes. In general, CVD processes have the advantage over PVD processes of good throwing power, while PVD processes have higher deposition rates than those in CVD processes.

Table 2 Selected parameters of basic PVD and CVD processes

Parameter	Physical vapor deposition		Chemical vapor deposition
	Evaporation	**Sputtering**	
Mechanism of production for depositing species	Thermal energy	Momentum transfer	Chemical reaction
Deposition rate	Can be very high (up to 7.5×10^5 Å/min)	Low except for pure metals (for example, copper, 10^4 Å/min)	Moderate (200-2500 Å /min)
Depositing specie	Atoms and ions	Atoms and ions	Atoms
Throwing power	Poor line-of-sight coverage except by gas scattering	Good, but nonuniform thickness distribution	Good
Metal deposition	Yes	Yes	Yes
Alloy deposition	Yes	Yes	Yes
Refractory compound deposition	Yes	Yes	Yes
Energy of deposited species	Low (1.6×10^{-20} to 8.0×10^{-20} J, or 0.1 to 0.5 eV)	Can be high (1.6×10^{-19} to 1.6×10^{-17} J, or 1 to 100 eV)	Can be high with plasma-assisted CVD
Bombardment of substrate/deposit	Not normally	Yes	Possible
Growth interface perturbation	Not normally	Yes	Yes (by rubbing)
Substrate heating (by external	Yes, normally	Not generally	Yes

As with all processes, the user is concerned with the process itself as well as the resulting microstructures and properties of the product. In order to understand various vapor deposition processes, one has to model them in terms of three steps (Ref 1):

- *Step 1:* creation of vapor phase specie (a material can be put into the vapor phase by evaporation, sputtering, laser ablation, electrolyte, chemicals, plasma spray, D-gun, gases, vapors, and so on)
- *Step 2:* transport from source to substrate (the vapor species are transported from the source to the substrate with or without collisions between atoms and molecules; during transport, some of the vapor species can be ionized by creating a plasma)
- *Step 3:* film growth on the substrate (this process involves the condensation of the vapor species onto the substrate and subsequent formation of the film by nucleation and growth processes; the nucleation and growth processes can be strongly influenced by bombardment of the growing film by ionic species, resulting in a change in microstructure, composition, impurities, and residual stress)

The degree of independent control of these three steps determines the versatility or flexibility of the deposition process. For example, these three steps can be independently controlled in PVD processes and therefore give a greater degree of flexibility in controlling the structure, properties, and deposition rate, whereas all of the three steps take place simultaneously at the substrate and cannot be independently controlled in the CVD process. Thus, if a process parameter such as the substrate temperature that governs the deposition rate in chemical vapor deposition is initially chosen, the user is limited to the resultant microstructure and properties obtained.

Physical Vapor Deposition Processes

The basic PVD processes fall into two general categories: (1) sputtering and (2) evaporation. PVD techniques are used for a wide variety of applications, from decorative uses to deposition of high-temperature superconducting films. The thickness of the deposits can vary from angstroms to millimeters. A very high deposition rate (25 μm/min, or 1000 μ in./min) has been achieved with the advent of electron beam (EB) heated sources. A very large number of inorganic materials (metals, alloys, compounds, and mixtures) as well as some organic materials can be deposited using PVD technologies. Ion plating is a hybrid PVD process because it is defined as an atomistic film deposition process in which the substrate surface and the depositing film is subjected to a flux of high-energy particles that is sufficient to cause changes in the interfacial region between the film and the substrate in addition to changes in the properties of the deposited film as compared to a nonbombarded film. These changes may be in the adhesion of the film to the substrate or in morphology, density, or stress. The source of the depositing species can be evaporation, sputtering, gases, or vapors.

Sputter Deposition

Sputtering is the phenomena of momentum transfer from an indirect energetic projectile to a solid or liquid target resulting in the ejection of surface atoms or molecules. In the sputter deposition process, the target (a source of coating material) and the substrate are placed in the vacuum chamber and evacuated to a pressure typically in the range of 13 to 0.013 mPa (0.1 to 10^{-4} mtorr). A diagram of the sputter coating process is shown in Fig. 1. The target (also called a cathode) is connected to a negative voltage supply, and the substrate generally faces the target. A glow discharge is initiated after an inert gas (usually argon gas) is introduced into the evacuated chamber. Typical working pressure of the argon is in the range of 3 to 20 Pa (20 to 150 mtorr). The sputter target erosion rate, *R,* in Å/min, is given by:

$$R = 62.3 \frac{JSM_A}{\rho} \qquad \textbf{(Eq 1)}$$

where *J* is the ion current density in mA/cm^2; *S* is the sputtering yield in atoms/ion; M_A is the atomic weight in grams; and ρ is the density of the target material in g/cm^3. Sputtering yield, *S,* assuming perpendicular ion incidence onto a target consisting of a random array of atoms, can be expressed as:

$$S = \text{constant} \cdot \frac{M_i\, M_t}{(M_i + M_t)} \cdot \frac{E_i}{U} \qquad \textbf{(Eq 2)}$$

where M_i is the mass of the incident atom, M_t is the mass of the target atom, E_i is the kinetic energy of the incident ion, and *U* is the heat of sublimation of the target material.

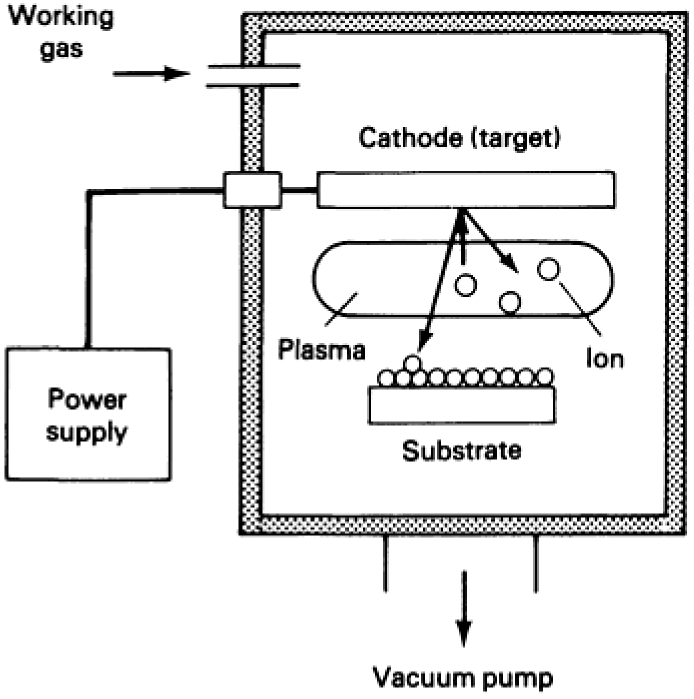


Fig. 1 Schematic showing primary components of a sputter deposition process system

Planar diode glow discharge deposition is the simplest sputtering system. This configuration consists of the cathode (the target) and anode facing each other. The substrates are placed on the anode. The target (which is usually water cooled) performs two functions during the process: one as the source of coating material, and the other as the electrode sustaining the glow discharge. The distance between the cathode and anode is usually about 50 to 100 mm (2 to 4 in.).

The discharge current increases with the applied voltage, thus increasing the sputtering rate. However, the discharge current does not increase linearly with the applied voltage as the voltage increases above 1.6×10^{-17} J (100 eV) because the ionization cross section decreases with increasing electron energy. The sputtering rate can be increased if the working gas pressure increases at a given voltage due to an increase in ion collection by the cathode (Ref 2). However, the deposition rate starts decreasing at high gas pressures due to gas scattering of the sputtered atoms.

The deposition rate is mainly determined by the power density at the target surface, the size of the erosion area, the source-to-substrate distance, the source material, and the working gas pressure. Some of these factors, such as pressure and power density, are interrelated. Therefore, the optimum operating condition is obtained by controlling the parameters to get the maximum power flux that can be applied to the target without causing cracking, sublimation, or melting. The maximum power limit can be increased if the cooling rate of the target is increased by designing the coolant flow channels properly and improving the thermal conductance between the target and the target backing plate.

Even though planar diode glow discharge sputter deposition techniques are widely used due to their simplicity and the relative ease of fabrication of targets for a wide range of materials, they have several disadvantages. Among the disadvantages of this deposition technique are the low deposition rate, the substrate heating due to the bombardment of high-energy particles, and the relatively small deposition surface areas.

Magnetron Deposition. By employing magnetic fields in the diode sputtering process, the ionization efficiency can be greatly increased. In the conventional planar diode process, ions are generated relatively far from the target and the probability that ions will lose their energy to the chamber walls is great. Furthermore, the number of primary electrons hitting the anode at high energies without experiencing collisions is increased as the pressure decreases, thus reducing ionization efficiency. These electron losses are not offset by impact-induced secondary emission. Therefore, ionization efficiencies are low and the discharge cannot be sustained in planar diodes at pressures <1.3 Pa (<10 mtorr).

In the magnetron sputter deposition process, an applied magnetic field parallel to the cathode surface forms electron traps and restricts the primary electron motion to the vicinity of the cathode. The magnetic field strength is in the range of a few hundred gauss: therefore, it can influence the plasma electrons but not the ions. The electrons trapped on a given field line can advance across the magnetic field to an anode or walls by making collisions (mostly with gas atoms). Therefore, their chances of being lost to the walls or anode without collisions are very small. Because of the higher efficiency of this ionization mechanism, the process can be operated at pressures of ~130 mPa (~1 mtorr) with high current densities at low voltages, thus providing high sputtering rates.

There are several configurations of magnetron sputter deposition technologies. Figure 2 shows key components of the cylindrical magnetron, the planar magnetron, the S-gun magnetron, and the unbalanced magnetron.

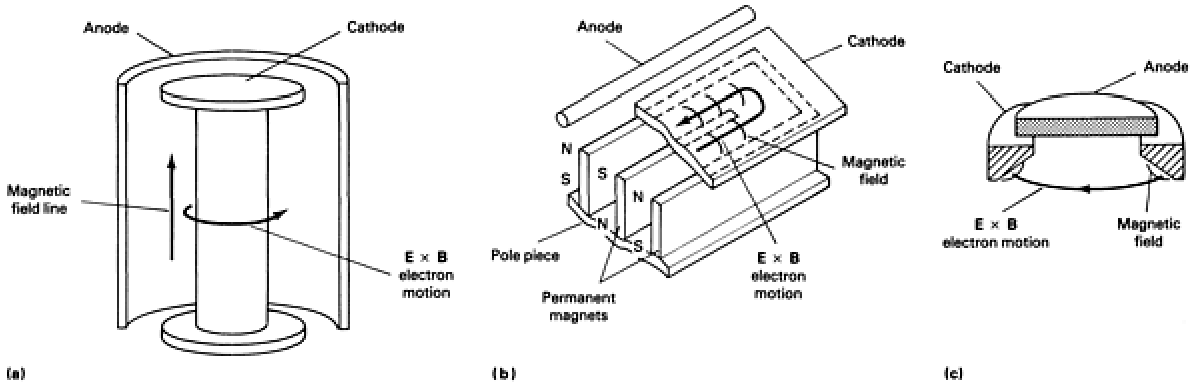


Fig. 2 Magnetron sputter deposition process units that use a variety of sources to supply the magnetic field. (a) Cylindrical magnetron source. (b) Planar magnetron source. (c) S-gun magnetron source

Cylindrical Magnetron. The cylindrical magnetron technique is very useful for preparing uniform coatings over large areas, because it employs long cathodes. Furthermore, the cylindrical-hollow magnetron technique is effective for coating objects with complex shapes. The cylindrical-post magnetron can be used to avoid substrate bombardment by energetic particles, thus preventing heating of the substrate.

Planar Magnetron. Metallic films and dielectric films can be deposited with high deposition rates using planar magnetron sputtering (as compared with those that can be achieved with diode sputtering.

S-Gun Magnetron. To deposit films on thermally sensitive substrate (such as electronic devices), the S-gun magnetron can be used because this technique allows good isolation of the substrate from the glow discharge plasma.

Unbalanced Magnetron. A very recent development is the unbalanced magnetron, in which some of the electrons from the target undergo collisions with gas atoms and form a plasma around the substrate from which ions are attracted to the substrate and bombard the growing film to produce changes in the structure and properties of the deposited film. In the normal balanced magnetron, the ion flux at the substrate is limited.

Magnetron Limitations. Even though the magnetron sputtering techniques have the advantages of high sputtering rates and low bombardment rates of energetic particles onto the substrate, the utilization of these techniques is impeded by the limitations in the choice of target materials and by the difficulties in fabricating the target. For example, if ferromagnetic materials are used as the sputtering target, they should be thin enough to ensure that the material is saturated by the magnetic field. Because high power is localized in a selected area in the magnetron sputtering process, targets should be prepared without voids or bubbles to avoid local melting and spitting.

Radio Frequency Deposition. The development of the radio frequency (rf) sputter deposition technique made it possible to deposit films from nonconducting sputtering targets, which cannot be sputtered by the direct current (dc) methods because of the charge accumulation on the target surface.

Most ions are almost immobile relative to electrons, which can follow the temporal variations in the applied potential at the typical rf frequencies used for sputtering (5 to 30 MHz). When the electrode is coupled to an rf generator, a negative

voltage is developed on the electrode due to the difference in mobility between the electrons and the ions. Because the insulating target electrode simulates a capacitor in the electrical circuit, there should be no dc component to the current flow. Therefore, the voltage on the electrode surface is required to be self-biased negatively to compensate for the difference in mobility of electrons and ions and to satisfy the condition of net zero current (averaged over each cycle). The magnitude of the resulting negative bias is almost the same as the zero-to-peak voltage of the rf signal. The period for the electrode to act as an anode is very short, and the electrode mostly acts as a cathode during the rf cycle. Therefore, the target can be expected to be sputtered as in the dc case. Significant numbers of ions are not accumulated on the target surface while the electrode acts as a cathode because of the high frequency employed in rf sputtering (normally 13.56 MHz).

Deposition can be performed at considerably lower pressures (typically, 0.7 to 2.0 Pa, or 5 to 15 mtorr) in rf sputtering as compared to the planar dc discharge. This is possible because electrons oscillating at high frequency can obtain sufficient energy to cause ionizing collisions and the number of electrons lost (without making collisions) can be reduced.

Radio frequency sputtering is widely used to deposit various kinds of materials (for example, conducting, semiconducting, and insulating coatings) despite the complexity of the rf power source. This technique can also be applied to magnetron sputtering sources.

Ion Beam Deposition. A relatively recent development called ion beam sputtering (Fig. 3) provides:

- Excellent adhesion
- A high-purity deposit resulting from the use of low operational pressures (~13 mPa, or 0.1 mtorr)
- Very low substrate hearing effects because the substrate is not in contact with the plasma

In this technique, an ion beam of high energy (hundreds to thousands of electron volts) extracted from the ion source is directed at a sputtering target of the desired material. An inert or reactive gas is used for the ion beam source. The substrate is suitably located to collect the sputtered species from the sputtering target (Fig. 3). Two primary ion sources are used in practical thin-film deposition: (1) the Kaufman source and (2) the duoplasmatron. No further discussion of these sources will be presented in this article; additional information is available in the book by Wilson and Brewer (Ref 3).

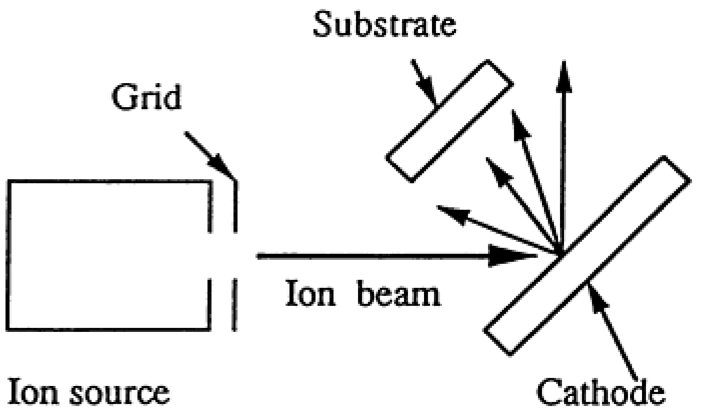


Fig. 3 Key components of an ion beam sputter deposition apparatus

Because the substrate can be isolated from the plasma generation source, ion beam deposition permits independent control over the substrate temperature, gas pressure, and the type of particle bombardment of the growing film. In addition, it is possible to control the energy and the target current density independently in this technique, whereas it can be done only by varying the working gas pressure in conventional glow discharge sputtering technology. The new unbalanced magnetron sputtering systems can also independently vary the ion flux and energy.

In general, the deposition rate in ion beam sputtering (50 nm/min, or 500 Å/min) is lower than that in conventional sputtering. The reason for this low rate is mainly due to the low beam current in the conventional (dual-grid) ion beam sputtering system. This low beam current can be greatly increased by using a single-grid system. Nishimura *et al.* (Ref 4) obtained a very high deposition rate (>90 nm/min, or 900 Å/min, for selected materials such as aluminum, copper, and so on) using single-grid ion beam sputtering. A small coverage of deposition area due to the small ion beam size (~10 mm, or 0.4 in.) is another drawback of the ion beam sputtering. The beam size can be increased to ~100 mm (~4 in.) by adaptation of space-type ion engine technology.

Evaporation Deposition

In evaporation processes, vapors are produced from a material located in a source that is heated by various methods (Fig. 4). The process uses an evaporation source to vaporize the desired material; the substrates are located at an appropriate distance facing the evaporation source. Resistance, induction, arc, electron beam, or lasers are the possible heat sources for evaporation. The substrate can be heated and/or biased to the desired potential using a dc/rf power supply. Evaporation is carried out in vacuum in a pressure range of 1.3×10^{-3} to 1.3×10^{-8} Pa (10^{-5} to 10^{-10} torr). In this pressure range, the mean free path (MFP) is very large (5 to 10^5 m, or 16 to 3×10^5 ft) as compared to the source-to-substrate distance. Hence, the evaporated atoms essentially undergo a collisionless line of sight transport prior to condensation on the substrate, leading to thickness buildup directly above the source that decreases steeply away from it. Planetary substrate holders are therefore used in some cases to even out the vapor flux on multiple substrates. In some cases, an appropriate gas such as argon at pressures of 0.7 to 30 Pa (5 to 200 mtorr) is introduced into the chamber to reduce the mean free path so that vapor species undergo multiple collisions during transport from the source to substrate, thus producing reasonably uniform thickness coatings on the substrate. This technique is called gas scattering evaporation or pressure plating (Ref 5).

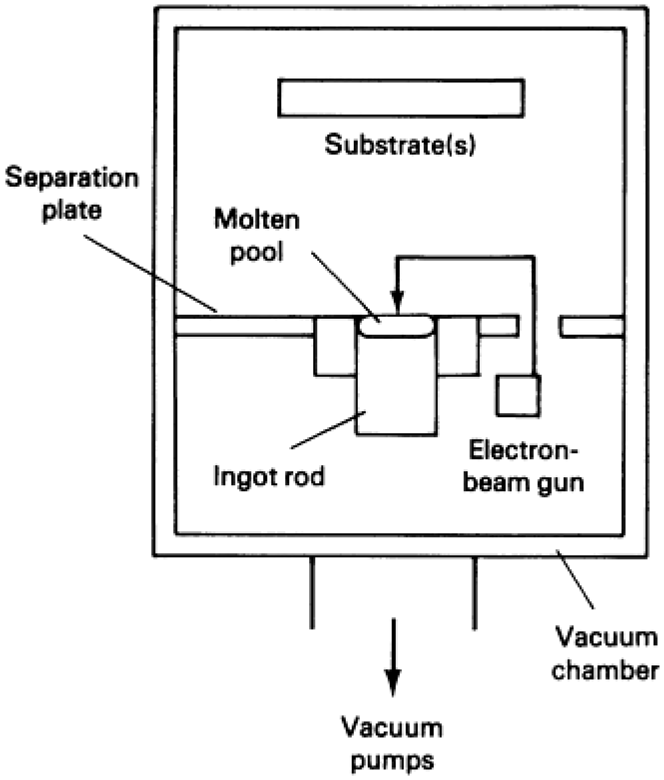


Fig. 4 Typical evaporation deposition process apparatus incorporating an electron beam as a heat source for evaporation

The transition of solids or liquids into the vapor phase is an atomistic phenomenon. It is based on thermodynamics and results in an understanding of evaporation rates, source-container reactions, and the accompanying effect of impurity introduction into the vapor state, changes in composition during alloy evaporation, and stability of compounds. An excellent detailed treatment of the thermodynamic and kinetic bases of evaporation processes is given by Glang (Ref 6).

The rate of evaporation is given by the well-known Hertz-Knudsen equation,

$$dN_e/A_e dt = \alpha_v (2\pi mkT)^{-1/2}(p^* - p) \qquad \textbf{(Eq 3)}$$

where α_v is the evaporation coefficient; $dN_e/A_e dt$ is the number of molecules evaporating from a surface area, A_e in time, dt; p^* is the equilibrium vapor pressure at the evaporant surface; p is the hydrostatic pressure acting on the surface; m is the molecular weight; k is Boltzmann's constant; and T is the absolute temperature. Evaporation coefficient, α_v, is very dependent on the cleanliness of the evaporant surface and can range from very low values for dirty surfaces to unity for clean surfaces.

For reasonable deposition rates (100 to 1000 nm/min, or 1000 to 10^5 Å/min) at a source-to-substrate distance of 200 mm (8 in.), the vapor pressure should be about 1.3 Pa (10^{-2} torr). The source temperature should be adjusted to give this value of the vapor pressure.

The directionality of evaporating molecules from an evaporation source is given by the cosine law. Holland (Ref 7) and others (Ref 8) have thoroughly discussed the theoretical distribution of vapor from a point, a wire, a small surface, an extended tip, and cylindrical and ring types of sources. For the ideal case of deposition from a clean, uniformly emitting point source onto a plane receiver, the rate of deposition varies as (cos θ)r^2 (Knudsen's cosine law), where r is the radial distance of the receiver from the source and $\theta$is the angle between the radial vector and the normal to the receiver direction.

PVD Techniques for Deposition of Metals, Alloys, and Compounds

The great versatility of the PVD processes is their ability to deposit a very large number of materials (for example, metals, alloys, semiconductors, superconductors, and polymers) and to fabricate composites of various types (for example, particulate, fibrous, or laminate).

Single-Element Specie Deposition

Single-element specie deposition can be carried out by evaporation or sputter deposition processes. The deposition rate depends on the process and the process parameters.

Alloy Deposition

Alloys consist of two or more components that have different vapor pressures and, hence, different evaporation rates. As a result, the composition of the vapor phase and the deposit are constantly varying. The following solutions have been used to obtain alloy films with stoichiometry close to the source composition by evaporation-based or sputter-based techniques.

- **Coevaporation or Cosputtering Using Multiple Sources.** This technique involves simulneous coevaporation of the constitutive elements of the alloy. The composition of the deposited film is controlled by adjusting the evaporation/sputtering rate of the respective elements. In elaborate systems, separate deposition rate monitors are used with appropriate feedback networks to control the deposition rate from each individual source independently. Near-stoichiometric films of many binary alloys have been successfully deposited using this technique. Dispersion-strengthened alloys such as Ni-ThO_2 have also been successfully deposited by this technique
- **Evaporation from a Single Source.** This technique involves evaporation of an alloy using a rod-fed electron beam source. Evaporation operates under steady-state conditions where the composition and volume of the liquid pool on the top of a solid rod are kept constant (Fig. 5). A detailed description of the process is given in Ref 9
- **Flash Evaporation.** In this process, pellets of the alloy are dropped onto a very hot strip and are vaporized completely, thus maintaining the composition of the alloy in the deposit. This technique works very well for elements with high vapor pressures
- **Sputter deposition from an alloy target**
- **Sputter deposition from a segmented target** where the segments consist of each of the two

components of the alloy and the ratio of the target sample area of each element is inversely proportional to the sputtering yields
- **Laser ablation from an alloy target**

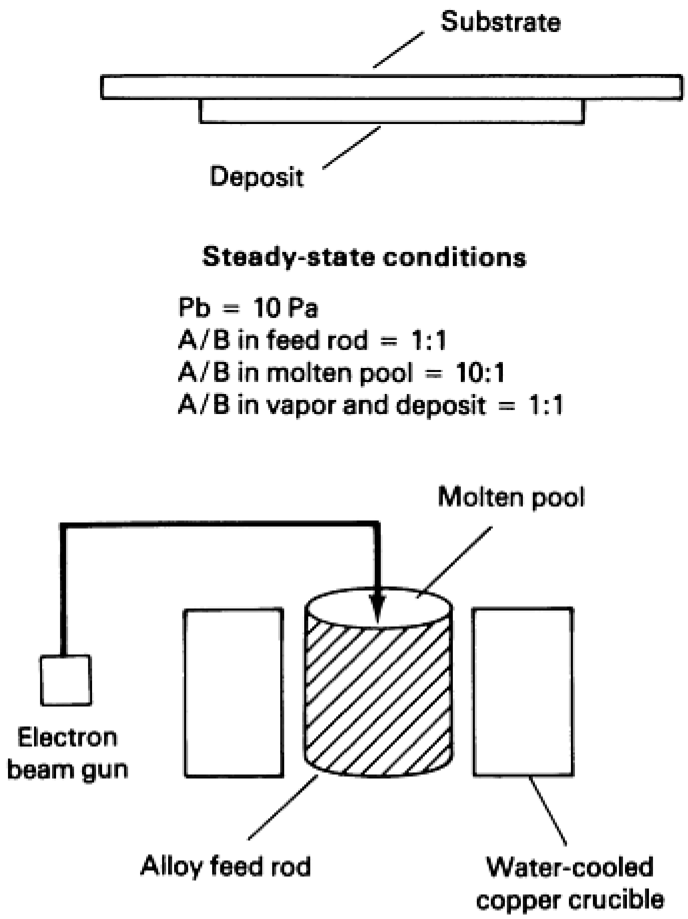


Fig. 5 Schematic showing direct evaporation of an alloy from a single-rod-fed electron beam source. Steady-state conditions: Vapor pressures, P^0_{B} = 10 P^0_{A}; in feed rod composition A_1B_1; molten pool composition $A_{10}B_1$; vapor and deposit composition A_1B_1

Deposition of Compounds

Deposition of compounds can be performed in two ways:

1. *Direct evaporation* using conventional heating methods or laser ablation, where the composition of the evaporant is the same as that of the compound that is to be deposited
2. *Reactive evaporation*, where the elements of the compound are evaporated and react with the gas to form the compound

Plasma-assisted reactive evaporation processes are often used because they activate the reactions leading to compound formation. Reactive evaporation and plasma-assisted reactive evaporation are discussed in the section "Reactive Evaporation Process" in this article.

Direct Evaporation. When a compound is heated, evaporation can occur with or without dissociation of the compound into fragments. A small number of compounds are evaporated without dissociation (for example, SiO_2, MgF_2, B_2O_3, GaF_2, and other Group IV divalent oxides).

Conventional Heating Methods. In the more general case, when a compound is evaporated, the material is not transformed to the vapor state as compound molecules but as fragments thereof. Subsequently, the fragments have to recombine on the substrate to reconstitute the compound. Therefore, the stoichiometry (anion-to-cation ratio) of the deposit depends on factors such as:

- Vaporization rate
- Ratios of the various molecular fragments in the vapor
- Impingement of other gases present in the environment on the film
- Surface mobility of the fragments (which in turn depends on their kinetic energy and substrate temperature) on the substrate
- Mean residence time of the fragments on the substrate
- Reaction rate of the fragments on the substrate to reconstitute the compound
- Impurities present on the substrate

For example, it was found that direct evaporation of Al_2O_3 resulted in a deposit that was deficient in oxygen. In another example, the deposit from direct evaporation of TiB_2 contained both the monoboride and diboride phases (Ref 10).

Laser ablation deposition techniques have attracted great interest in recent years for the synthesis of semiconducting and insulating thin films. Very recently, these techniques have been applied successfully for the deposition of high-critical-temperature (high-T_c) superconducting films (Ref 11, 12).

In the laser ablation technique, material is vaporized and ejected from the surface of a target as it is irradiated by a laser beam. Films are formed by condensing material ablated from the target onto a solid substrate. Absorption characteristics of the material to be evaporated determine the laser wavelength to be used. To obtain the high-power density required in many cases, pulsed laser beams are generally employed. Pulse width, repetition rate, and pulse intensity are selected for specific applications.

Although laser ablation is an attractive approach for the synthesis of high-purity metal, alloy, and compound films, it suffers from the following limitations:

- Complex transmitting and focusing systems need to be employed to direct the beam from the laser located outside the vacuum chamber onto the evaporant placed inside the system. This involves special designs and increases the cost of the setup. In addition, a window material that efficiently transmits the wavelength band of the laser must be found and mounted in such a way that it is not rapidly covered up by the evaporant flux
- It is not always possible to find a laser with a wavelength that is compatible with the absorption characteristics of the material to be evaporated
- Energy conversion efficiency is very low
- The size of the deposited film is small (10 to 20 mm, or 0.4 to 0.8 in., diameter), resulting from the small size of the laser impact spot

Reactive PVD processes are those in which at least one of the elements of the coating is introduced in the gas phase. Various compounds have been synthesized from metal targets with reactive gases:

- Oxides with air, O_2, or H_2O
- Nitrides with N_2 or NH_3
- Oxynitrides with $O_2 + N_2$
- Sulfides with H_2S

There are several advantages to these processes. For example, various kinds of compounds can be prepared using relatively easy-to-fabricate metallic targets, insulating compound films can be prepared using direct current (dc) power supplies (reactive sputtering), and graded composition films can be formed.

Reactive Sputtering (RS). In the RS process, sputter deposition is performed in the presence of reactive species in the gas phase. Reactions can occur on the cathode surface, at the substrate, and in the gas phase. However, reactions in the vapor phase are precluded by considerations of momentum and energy conservation unless the process is performed at high pressures to allow multiple-body collisions in the gas phase.

The target is a nominally pure metal. The compound film is synthesized by sputtering in a pure reactive gas or mixture of inert gas and reactive gas. Usually, the inert-reactive gas mixture is preferred because of sputtering rate considerations. A compound target also can be used in the RS technique. In this case, the target is chemically decomposed by inert gas ion bombardment. It is usually necessary to add the reactive gas to compensate for the loss of the reactive component by dissociation.

The main problem in reactive sputtering is target poisoning. As the reactive gas partial pressure increases, the rate of compound formation exceeds the removal rate of compounds on the target surface, resulting in a decrease in deposition rate due to the low sputtering yield of the compound formed on the cathode surface and to the fact that compounds have higher secondary electron emission yield than pure metals. The increase in secondary electron emission results in a reduction in both the discharge voltage and ion component in the cathode current at constant voltage. In other words, more of the energy of incoming ions is consumed to produce and accelerate secondary electrons. Various solutions have been found to reduce the effect of target poisoning (Ref 13).

Reactive Evaporation Process. The difficulties involved in direct evaporation processes due to fragmentation of the vaporized compounds are overcome in reactive evaporation, where a metal is evaporated in the presence of the reactive gas. The compound is formed by reaction of the evaporating metal species with the molecules of the reactive gas. Even though this technique has been extensively used to deposit a variety of oxide films for optical applications, it is generally observed that the films are deficient in oxygen. It is also observed in some cases (for example, the synthesis of carbide films) that the deposition rate becomes a limiting factor governing the growth of the films. In such cases, stoichiometric titanium carbide films could be deposited only at very low rates (~0.15 nm/s, or 1.5 Å/s). This deposition rate limitation is due to the reaction kinetics of the compound formation in this process. The presence of a plasma in the activated reactive evaporation (ARE) process influences the reaction kinetics by providing activation energy to the reactive species, thereby making it possible to synthesize compound films at considerably higher rates (Ref 14) and lower temperatures.

Activated Reactive Evaporation. The ARE process generally involves evaporation of a metal or an alloy in the presence of the plasma of a reactive gas (Ref 15, 16). For example, titanium carbide and titanium nitride coatings are deposited with this process by evaporating titanium in the presence of C_2H_2 and N_2 plasmas, respectively. The two basic variants of the ARE process (one version incorporates an electron beam to evaporate the target material, whereas the other version utilizes resistance elements to provide the heat required to evaporate the target material) are shown in Fig. 6. Detailed information on other modifications of the ARE process is available in the article by Bunshah and Deshpandey (Ref 17).

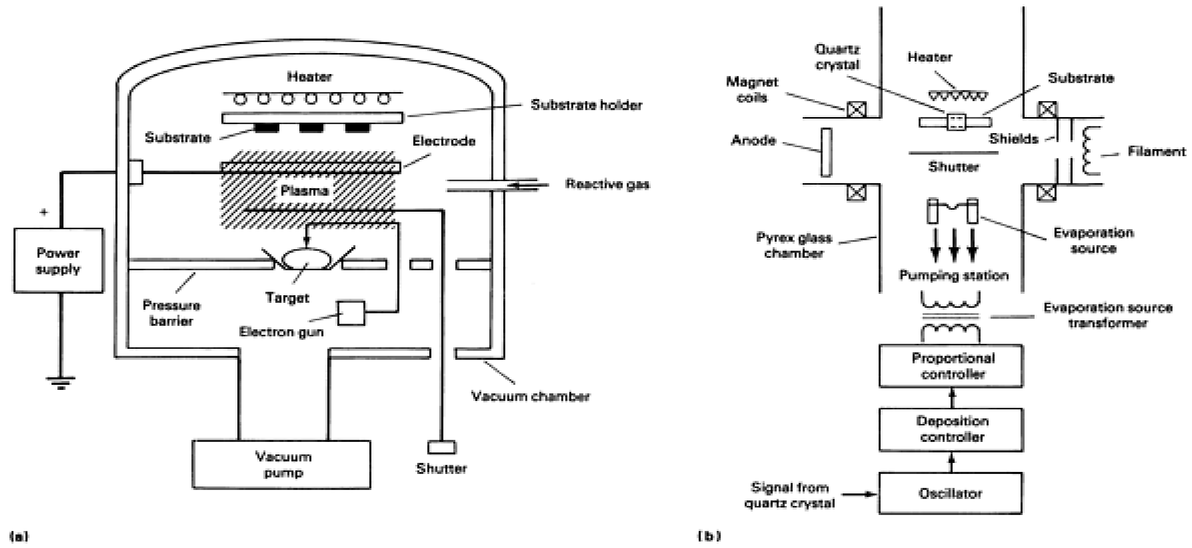


Fig. 6 Activated reactive evaporation (ARE) process equipment incorporating different heat sources to evaporate target materials. (a) Electron beam source. (b) Resistance element heating source

The role of the plasma is two-fold:

1. To enhance the reactions that are necessary for deposition of compound films
2. To modify the growth kinetics and, hence, the structure/morphology of the deposits

Hybrid PVD Processes. Ion plating and reactive ion plating (RIP) are two deposition processes classified as hybrid PVD processes.

Ion plating is a hybrid process (Ref 18) because the coating material is vaporized by thermal energy (that is, evaporation) or momentum transfer (that is, sputtering) or supplied as a vapor (very similar to CVD processes). In this technique, the vaporized (or supplied) coating materials pass through a gaseous glow discharge on their way to the substrate, thus ionizing some typically 1% of the vaporized atoms.

The glow discharge is produced by biasing the substrate to a high negative potential (-2 to -5 kV) and admitting a gas (usually argon) at a pressure of 0.7 to 30 Pa (5 to 200 mtorr) into the chamber (Fig. 7). In this simple mode, which is known as diode ion plating, the substrate is bombarded by high-energy gas ions that sputter off the material present on the surface. This results in a constant cleaning of the substrate (that is, a removal of surface impurities by sputtering), which is desirable for producing better adhesion and lower impurity content. The ion bombardment also causes a modification in the microstructure and residual stress in the deposit. On the other hand, it produces the undesirable effects of decreasing the deposition rate because some of the deposit is sputtered off; it also causes a considerable (and often undesired for microelectronic applications) heating of the substrate by the intense ion bombardment. The latter problem can be alleviated by using the supported discharge ion-plating process, in which the substrate is no longer at the high negative potential because the electrons necessary for supporting the discharge come from an auxiliary heated tungsten filament. The high gas pressure used during deposition causes a reasonably uniform deposition on all surfaces due to gas scattering.

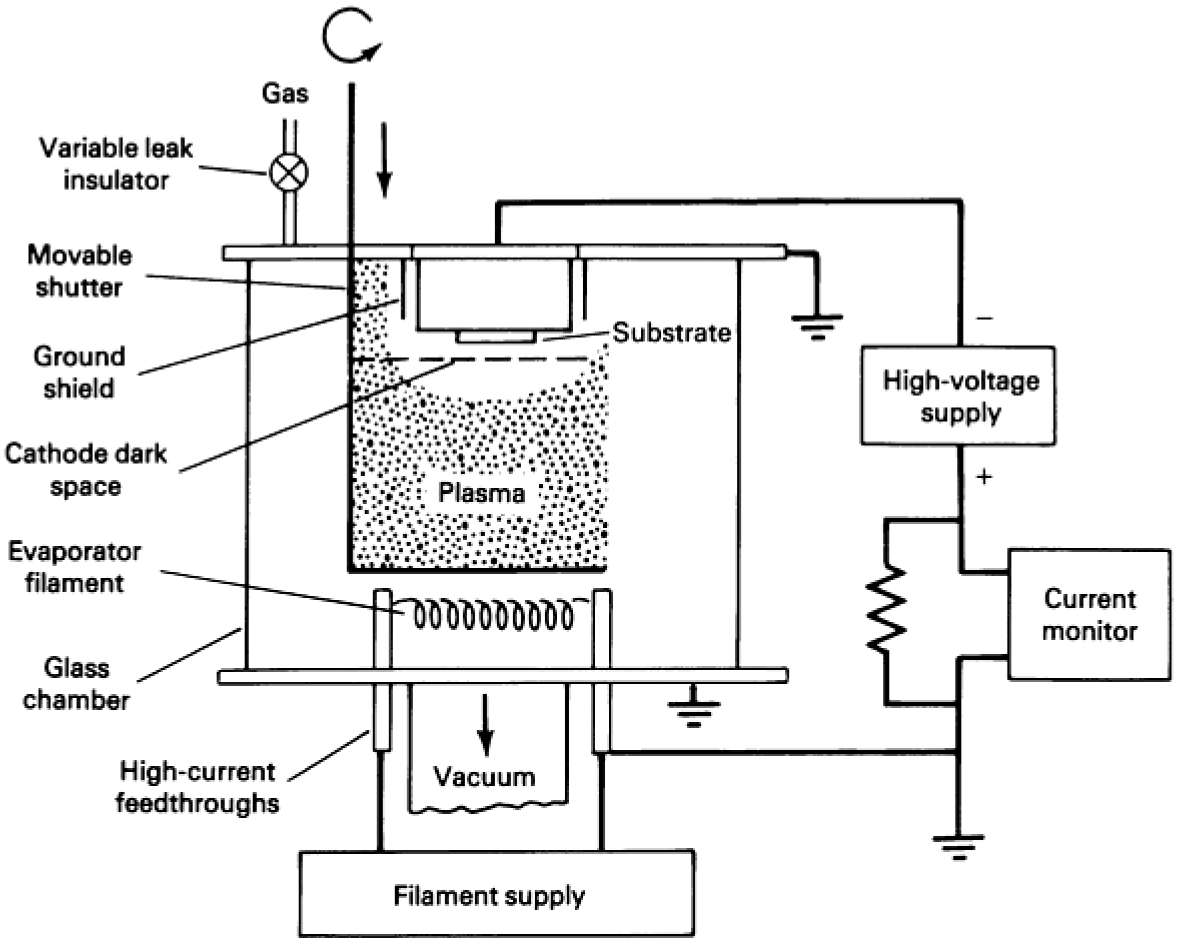


Fig. 7 Typical ion-plating process system that uses a dc diode discharge to ionize the vaporized atoms

Reactive ion plating is very similar to the reactive evaporation process in that the metal atoms and reactive gases, aided by the presence of a plasma, react to form a compound. Because the partial pressure of the gases in reactive ion plating is much higher (>1.3 Pa, or 10^{-2} torr) than in the ARE process (>0.013 Pa, or 10^{-4} torr), the deposits can become porous or sooty. The plasma cannot be supported at lower pressures in the simple diode ion-plating process. Therefore, Kobayashi and Doi (Ref 19) introduced an auxiliary electrode biased to a positive low voltage (as originally conceived for the ARE process) to initiate and sustain the plasma at lower pressure (~0.13 Pa, or 1 mtorr) (Fig. 8). This is no different from the ARE process with a negative bias on the substrate reported much earlier by Bunshah (Ref 20) and designated by him as the biased ARE (BARE) process.

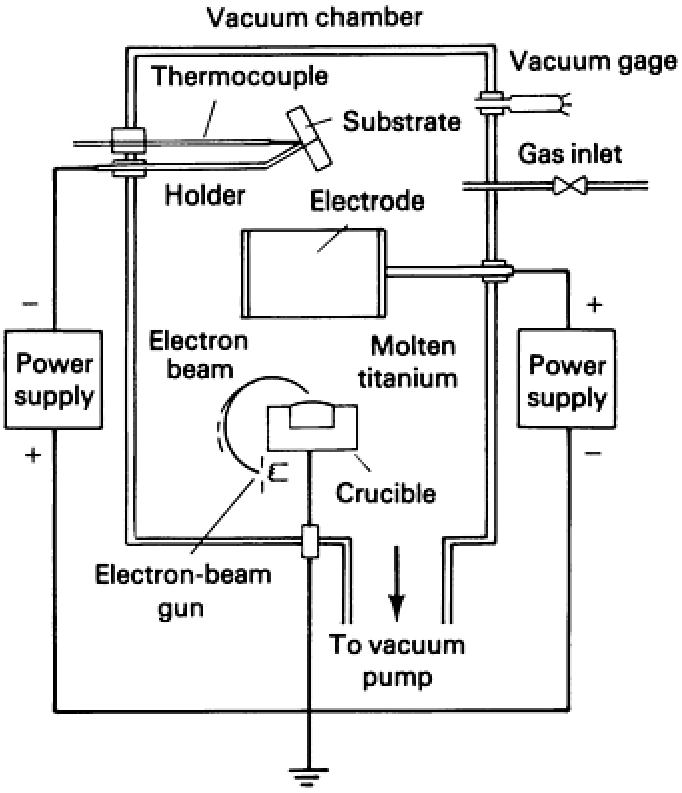


Fig. 8 Schematic of a reactive ion plating (RIP) system that incorporates an auxiliary electrode biased to a positive low voltage to generate plasma at relatively low pressure

Another variation of reactive ion plating using a triode configuration (Ref 21) involves injection of electrons into the reaction zone between the electron-beam-heated evaporation source and the negatively biased substrate from a heated tungsten filament mounted transversely to the metal vapor path. These low-energy electrons are pulled across the reaction zone by a positively biased anode located opposite the cathode. The arrangement is very similar to that shown in Fig. 8 except that it uses an electron-beam-heated (EB-heated) evaporation source; it is also very similar to triode sputtering. The EB-heated evaporation source adds versatility as well as complexity to the process through the addition of another process variable. Murayama (Ref 22) used an EB-heated source with a negatively biased substrate and rf activation of the reactants by means of a coil electrode of aluminum wire in the reaction zone to deposit oxide and nitride films.

Chemical Vapor Deposition Processes

Chemical vapor deposition processes are widely used in industry because of their versatility for depositing a very large variety of elements and compounds covering a wide range of materials from amorphous deposits to epitaxial layers having a high degree of perfection and purity.

Chemical vapor deposition can be defined as a process in which the gaseous chemical reactants are transported to the reaction chamber, activated thermally (conventional CVD) or by other means (plasma-assisted CVD or laser-induced CVD) in the vicinity of the substrate, and made to react to form a solid deposit on the substrate surface. It is possible to deposit films of uniform thickness and low porosity even on substrates of complicated shape with this process. If a metal-organic or organic-metallic precursor species is used in a CVD process, the process is designated MOCVD or OMCVD.

A major area for utility of chemical vapor deposition is in microelectronics applications (for example, gate-insulating layers, passivation layers, oxidation barriers, doped epitaxial layers of silicon, polycrystalline silicon layers, and so on). CVD techniques are also extensively employed to deposit protective coatings for a variety of operating environments where protection is required against the effects of wear, erosion, and high-temperature oxidation.

Conventional CVD (Thermal CVD) Processes

In conventional CVD (CCVD), the gaseous reactants are activated thermally in the vicinity of the heated substrate and react to form a film on the substrate. This technique is referred to simply as chemical vapor deposition in many cases (Fig. 9).

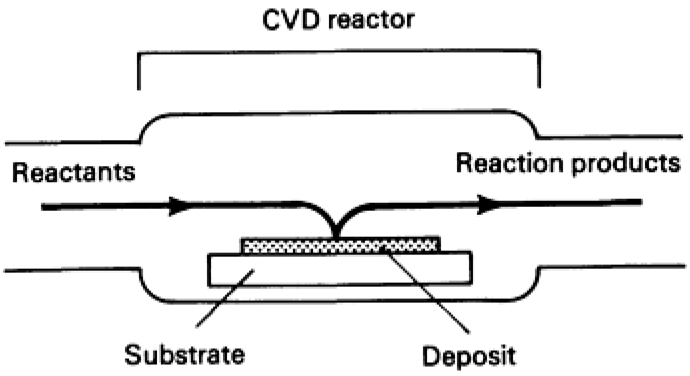


Fig. 9 Key components in a conventional chemical vapor deposition (CCVD) system

Heating sources used in CVD processes include:

- *Hot plate:* The substrate is in direct contact with a hot plate that is either resistively or inductively heated
- *Radiant heat:* The substrate is heated by a thermal radiation technique or optical technique (tungsten filament lamp or laser)
- *Heating of a conductive substrate:* Conductive substrates can be heated resistively or by rf induction

Classification of CVD Reactions. CVD reactions fall into four general categories: (1) thermal decomposition reactions, (2) reduction reactions, (3) displacement reactions, and (4) disproportionation reactions.

Thermal Decomposition Reactions (Pyrolytic Reactions). This type of reaction is characterized by:

$$AX(g) \rightarrow A(s) + X(g) \qquad \textbf{(Eq 4)}$$

where *AX* is a gaseous compound, *A* is a solid material, and *X* is a gaseous reaction product. Some examples of these reactions are:

$$B_2H_6(g) \rightarrow B(s) + 3H_2(g) \qquad \textbf{(Eq 5)}$$

$$SiH_4(g) \rightarrow Si(s) + 2H_2(g) \qquad \textbf{(Eq 6)}$$

$$W(CO)_6(g) \rightarrow W(s) + 6CO(g) \qquad \textbf{(Eq 7)}$$

Reduction Reactions. In these reactions, a gaseous compound is reduced by a reducing agent (usually hydrogen):

$$2AX(g) + H_2(g) \rightarrow 2A(s) + 2HX(g) \qquad \textbf{(Eq 8)}$$

Typical reduction reactions include:

$$2BCl_3(g) + 3H_2(g) \rightarrow 2B(s) + 6HCl(g) \qquad \textbf{(Eq 9)}$$

$$SiCl_4(g) + 2H_2(g) \rightarrow Si(s) + 4HCl(g) \qquad \textbf{(Eq 10)}$$

Displacement Reactions. These reactions are also known as exchange reactions. In the molecule *AX*, *X* is replaced by another element, *B*.

$$AX(g) + B(g) \rightarrow AB(s) + X(g) \qquad \textbf{(Eq 11)}$$

Some representative reactions include:

$$Zn(g) + H_2S(g) \rightarrow ZnS(s) + H_2(g) \qquad \textbf{(Eq 12)}$$

$$SiCl_4(g) + CH_4 \rightarrow SiC(s) + 4HCl(g) \qquad \textbf{(Eq 13)}$$

$$CrCl_2(g) + 2Fe(s) \rightarrow \text{Cr-Fe alloy} + FeCl_2(g) \qquad \textbf{(Eq 14)}$$

Disproportionation Reactions. In these reactions the oxidation number of an element both increases and decreases through the formation of two new species. Some typical examples of disproportionation reactions are:

$$2GeI_2(g) \rightarrow Ge(s) + GeI_4(g) \qquad \textbf{(Eq 15)}$$

$$2TiCl_2(g) \rightarrow Ti(s) + TiCl_4(g) \qquad \textbf{(Eq 16)}$$

Complex CVD Reactions. Several types of reactions can be involved simultaneously in some CVD coating processes. An example of these reactions is:

$$2AlCl_3(g) + 3CO_2(g) + 3H_2(g) \rightarrow Al_2O_3(s) + 3CO_2(g) + 6HCl(g) \qquad \textbf{(Eq 17)}$$

In this reaction, water produced from:

$$CO_2(g) + H_2(g) \rightarrow CO(g) + H_2O(g) \qquad \textbf{(Eq 18)}$$

is used to form Al_2O_3 by the following reaction:

$$2AlCl_3(g) + 3H_2O(g) \rightarrow Al_2O_3(s) + 6HCl(g) \qquad \textbf{(Eq 19)}$$

Rate-Limiting Steps. The sequence of events in a CVD process is as follows (Ref 23):

1. Diffusion of reactants to the surface
2. Adsorption of reactants at the surface
3. Surface events (for example, chemical reaction, surface motion, lattice incorporation, and so on)
4. Desorption of products from the surface
5. Diffusion of products away from the surface

Among these five steps, the slowest one is the rate-determining step.

The rate-limiting step is mainly determined by the process parameters. The most important rate-limiting steps in the CVD process are mass transport control and surface kinetics control. The latter produces uniform deposits on complex-shaped substrates.

Reactors. Two kinds of reactors are used most frequently in the CVD processes: hot-wall reactors (Fig. 10a) and cold-wall reactors (Fig. 10b).

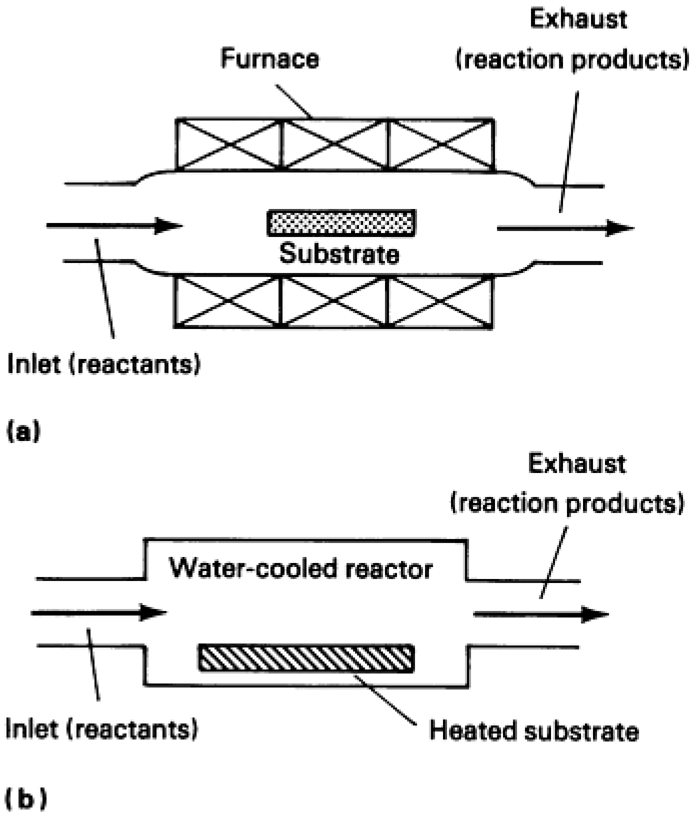


Fig. 10 Two types of chemical vapor deposition (CVD) reactors. (a) Hot-wall reactor. (b) Cold-wall reactor

Hot-Wall Reactors. In the hot-wall reactor, the reactor tube is surrounded by a tube furnace making the temperature of the substrate equal to the temperature of the reactor wall. A large number of substrates can be coated in this type of reactor. A major drawback of this type of reactor is deposition on the reactor wall and possible contamination in the system from chemical reactions between the reactor wall and the vapor due to the high temperature of the reactor wall. Therefore, the hot-wall reactor is ideal for use where the reaction is exothermic because the high wall temperature prevents undesirable deposition on the reactor walls. Groups III-V and II-VI type semiconductors have been successfully prepared in the hot-wall reactor.

Cold-Wall Reactor. In the cold-wall reactor, only the susceptor where the substrates are placed is intentionally heated by rf induction or high-radiation lamps. This type of reactor is predominantly used for the deposition reaction, which is endothermic (for example, silicon deposition from the halides or the hydride). Because the substrates have a higher temperature than the reactor wall, the reaction will proceed most readily on the hot surface of the substrate. In this type of reactor, contamination due to the interaction between the reactor wall of the vapor can be greatly reduced. Very frequently, the walls are water-cooled to further prevent deposition on the wall or reactions between the walls and the vapor.

Low-Pressure Chemical Vapor Deposition

Low-pressure CVD (LPCVD) technology is widely used in the semiconductor industry because it offers several advantages over the conventional atmospheric-pressure CVD technique. By operating at lower pressure to increase diffusivity in the gas phase and increasingly subjecting the system to surface kinetics control, the uniformity of deposition

is enhanced. The mass transfer rate and the surface reaction rate are generally of the same order of magnitude at normal atmosphere pressure (CCVD), while the mass transfer rate is much higher than the surface reaction rate at lower pressure (that is, 70 to 130 Pa, or 0.5 to 1 torr) using LPCVD and thereby the rate-determining step is the surface reaction. The transfer rates of gaseous reactants and reaction products are inversely proportional to pressure. If the pressure is reduced from 1.0×10^5 Pa (760 torr) to 70 to 130 Pa (0.5 to 1 torr), diffusivity increases by a factor of 1000, which is only partially offset by the increase in thickness of the boundary layer (by the square root of pressure) (Ref 24). At this low pressure, mass transfer cannot be the rate-limiting step and the deposition rate is mainly controlled by surface reactions, resulting in uniform film thickness and properties over extended surfaces with better step coverage and conformality, and good structural integrity with fewer pinholes.

Another advantage of low-pressure chemical vapor deposition occurs because the mean free path is very large at lower pressures; therefore, wafers can be stacked on edged instead of lying flat, and thus a large number of wafers can be loaded and deposited in the same run (Fig. 11).

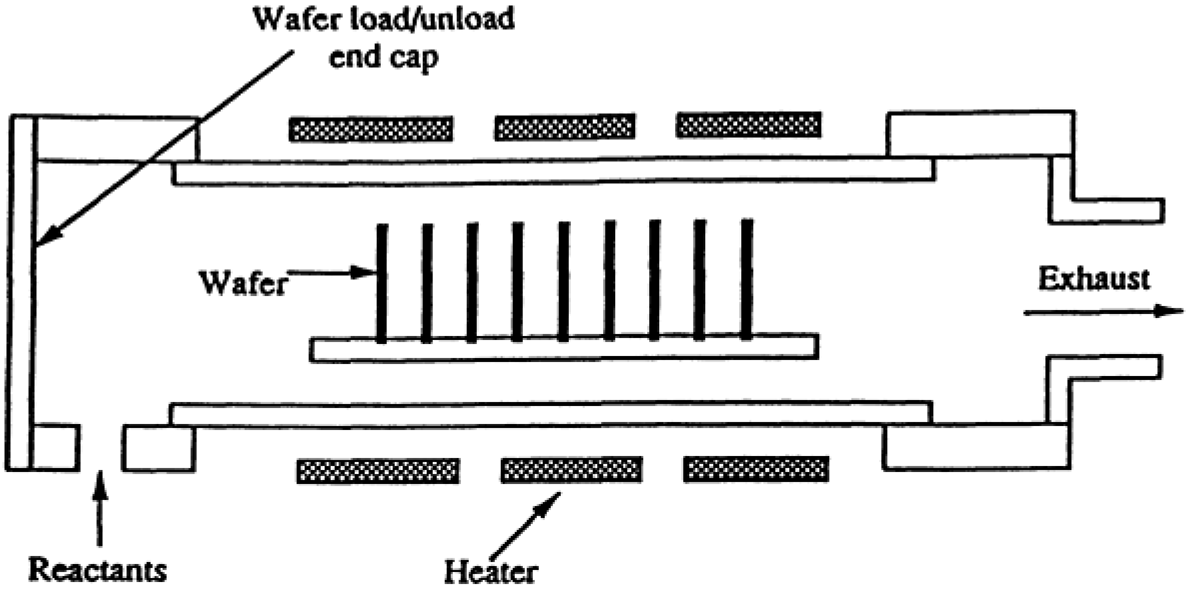


Fig. 11 Typical low-pressure chemical vapor deposition (LPCVD) reactor used in semiconductor industry applications to coat silicon wafers

Polycrystalline silicon films from SiH_4, Si_3N_4 films from SiH_2Cl_2 and NH_3, and SiO_2 films from SiH_2Cl_2 and N_2O have been successfully prepared using this technique (Ref 25, 26).

Plasma-Assisted Chemical Vapor Deposition

Plasma-assisted chemical vapor deposition (PACVD) can be defined as a process in which the constituents of the vapor phase react to form a solid film when assisted by an electric discharge. In the PACVD technique, the gas molecules are mainly dissociated by electron impact to generate very reactive neutral, radical, and ion species. These reactive species arrive on a surface and react with each other via an ionic or free radical mechanism in the film-forming process. Because the gas molecules are activated by the energetic electrons instead of thermal energy, the reaction temperature can easily be reduced. Films can be deposited at temperatures typically <300 °C (<570 °F). Furthermore, the inherent limitations of conventional thermodynamics and of chemical availability in conventional chemical vapor deposition are eliminated in plasma activation (Ref 27) due to the nonequilibrium nature of the glow discharge plasma. Key components of a microwave plasma CVD apparatus are shown in Fig. 12.

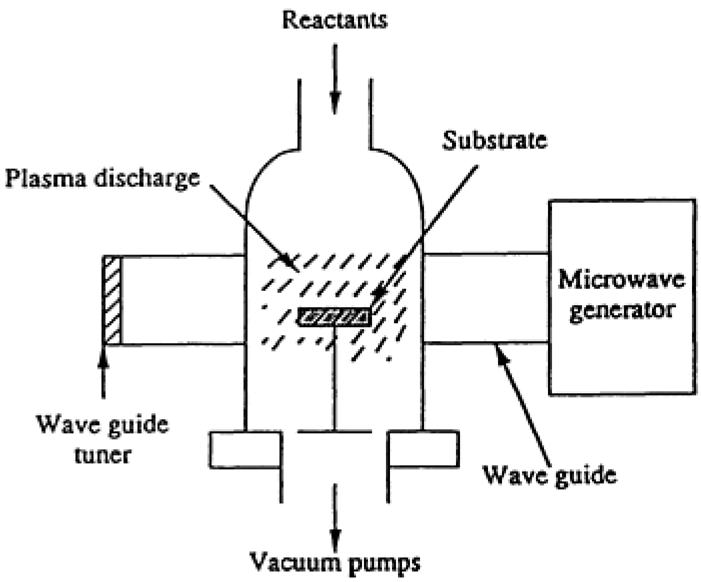


Fig. 12 Schematic showing key components of a microwave plasma chemical vapor deposition apparatus

Neutral radicals are believed to be the major deposition agents among the reactive species generated in the plasma due to the following two effects (Ref 24):

- Dissociation energy is usually lower than the ionization energy for many gas molecules; therefore, the generation rate for radicals is generally greater than that for ions
- Positive ions may drift toward any surface and recombine with electrons, whereas neutral radicals may have more chance to stay in the plasma and have a longer lifetime

The number of neutral radicals is determined by the following parameters:

- Glow discharge power
- Electrode spacing
- Gas collision mean free path
- Ion diffusivity

These variables determine the electron temperature (energy states of the electrons in the plasma) and thereby the radical generation rate. The radical generation rate is roughly proportional to $e^{-(\Delta E)/kT_e}$, where ΔE is the dissociation energy of gas molecules; T_e is the electron temperature; and k is the Boltzmann constant. This relation indicates that the number of gas radicals that have higher dissociation energy will change more with the electron temperature than those of the species with lower dissociation energy (Ref 26). This change in the valued of the radical ratio is the primary factor determining the film stoichiometry.

Ion Bombardment Energy. Even though ions do not contribute greatly to the film-forming process (Ref 27, 28), the ions impinging on the film during growth significantly affect the physical properties of the film (Ref 29). The ionic bombardment energy is determined by several factors (for example, discharge power, gas type and pressure, target bias, and radio frequency). At lower pressure (<7 Pa; or 0.05 torr) and low frequency $\ll$1 MHz), the ions can bombard onto the surface with the full energy available to them because the ions can follow the rf field and do not experience appreciable scattering. The ionic bombardment energy is an average plasma potential as the pressure and the frequency increase.

Substrate Temperature. The substrate temperature still plays an important role in the PACVD process, although the activation energy for the chemical reaction is provided mainly by the glow discharge. During film growth, the absorbed radical has to diffuse to a stable site to become part of the growing film. This radical mobility on the surface of the substrate is strongly affected by substrate temperature. The radicals on the surface obtain more energy and diffuse to the stable site easily at a high substrate temperature to produce a denser film. At low temperatures, however, the diffusion of adatoms (that is, atoms adsorbed on a surface so that they will migrate over the surface) on the surface is extremely retarded and therefore the film has more defects and a lower density. Furthermore, the stoichiometry of the film can be affected by substrate temperature. In the case of PACVD silicon nitride, the content of hydrogen in the grown film can be reduced greatly by increasing the substrate temperature.

Advantages of PACVD over CCVD Processes. There are several advantages of PACVD processes over CCVD processes:

- Ability to deposit films at much lower temperature
- Can be used to prepare the heavily hydrogenated amorphous silicon films for solar cells
- Good adhesion and bond strength of the films to their substrates
- Higher deposition rates than the CCVD technique

Limitations of PACVD Processes. The present limitations of PACVD processes include:

- More numerous defects in the films and a lower density of the film compared to a high-temperature deposited film
- Difficulties in deposition of pure materials
- Extreme difficulty in controlling stoichiometry

Advanced CVD Techniques

The two latest state-of-the-art CVD technologies are known as hot-filament chemical vapor deposition and laser-induced chemical vapor deposition (LCVD).

Hot-Filament CVD. Yasui *et al.* (Ref 30) prepared high-quality silicon nitride films with a low concentration of hydrogen at high deposition rates by the hot-filament CVD technique. In this technique, the hot filament (tungsten filament), which is heated to very high temperatures (for example, 2400 °C, or 4350 °F) is placed close (typically, within 80 mm, or 3.1 in.) to the substrate. Yasui deposited films at a deposition rate that was one order of magnitude higher with a smaller activation energy of the growth rate than that of CCVD technique using the same gases. The same techniques are now used to grow diamond films.

Laser-Induced CVD. Because the substrate is bombarded by charged and energetic particles in addition to high-energy radiation in the PACVD process, this technique can introduce deleterious effects for radiation-sensitive electronic devices. Recently, laser-induced CVD processes have been attracting widespread attention. In the LCVD process, the reaction energy is selectively provided by photons.

There are two types of LCVD processes: pyrolytic LCVD and photolytic LCVD.

Pyrolytic LCVD. In pyrolytic LCVD, polyatomic gas molecules are dissociated near a gas-substrate interface by localized heating of the substrate, which is exposed to the laser beam. This technique is limited by the choice of the laser/source gas/substrate system. The gas vapor sources are required to be relatively transparent at the exciting laser wavelength, and the substrates must be strongly absorbing (Ref 29).

Photolytic LCVD. In photolytic LCVD, a molecule near the substrate is decomposed by means of a photochemical reaction. Specific chemical bonds in polyatomic molecules can be broken selectively through the choice of the laser wavelength (Ref 30). In this technique, photodissociation of the vapor source fixes the maximum allowable wavelength of the laser because only radiation that is absorbed by the reactants can lead to a photochemical reaction. LCVD techniques have been successfully employed to synthesize polycrystalline silicon and germanium (Ref 31), aluminum oxide (Ref 32), and silicon (Ref 33).

Materials Deposited by PVD and CVD Techniques

In addition to the deposition of simple element species (for example, refractory metals by chemical vapor deposition for large-scale integrated circuits in addition to chemical applications, aluminum for metallization and decorative coatings, and so on), a number of more complex materials are deposited using PVD and CVD techniques, particularly the plasma-assisted versions listed in Table 3 (Ref 34).

Table 3 Compounds and synthesized deposition rates obtained with commonly used plasma-assisted deposition processes

Material		Plasma-assisted physical vapor deposition		
Compound	**Examples**	**Activated reactive evaporation, /min**	**Reactive sputtering, /min**	**Plasma-assisted chemical vapor deposition, /min**
Common				
Carbides	TiC, SiC, HfC, ZrC, VC	2000-3000	400-1000	150-400
Nitrides	TiN, HfN, ZrN	2000-3000	300-1000	60-150
Oxides	TiO_2, ZrO_2, Al_2O_3, SiO_2	1000-2000	200-800	200-300
Sulfides	TiS_2, MoS_2, MoS_3	1000-2000	. . .	. . .
Exotic				
Superconducting materials	Nb_3Ge, $CuMo_6S_8$	1000-1500	. . .	. . .
Photovoltaic materials	a-SiH_x, $CuInS_2$, $CuInSe_2$[(a)]	1500-2000	50-200	50-100
Optoelectronic materials	Indium tin oxide, ZnO	500-1000	50-75	50-100
Cubic boron nitride		1000-1500	. . .	. . .
Diamond		1-2[(b)]	. . .	1000-3000[(c)]
a-Carbon		1-2[(b)]	300	200

(a) a, amorphous.

(b) m/h.

(c) /h

Wear Applications

This subject is extensively treated in a book by B. Bhusan and B.K. Gupta (Ref 35). Soft coatings are extensively used for lubrication or lowered coefficient of friction. On the other hand, hard coatings are essential for good wear resistance with the accompanying low coefficient of friction. Such coatings are primarily carbides, nitrides, oxides, or borides of simple elements or alloys deposited by PVD and CVD techniques. These include relatively simple compositions such as TiN, TiC, and Al_2O_3; more complex compositions such as (TiAl)N and Ti(C,N); and multilayer coatings, such as TiC/Al_2O_3/TiN. It must be emphasized that the substrate must be hard and must also be resistant to elevated temperatures for many applications of hard coatings. The type of wear (for example, adhesive, fretting, abrasive, impact, chemical, and various combinations thereof) encountered must also be considered. The coating has to be tailored to the specific application. For example, the hardest substance known to man, diamond, is not useful for machining ferrous alloys, titanium alloys, or zirconium alloys, due to the chemical wear that occurs between diamond and these metals at high temperatures. Diamond (carbon) dissolves in these metals at high temperatures. Nevertheless, one of the exciting recent developments in hard coatings is the commercial availability of diamond-coated tools to machine aluminum alloys, plastics, and composites.

It is difficult, if not impossible, to provide general guidelines that relate the coefficient of friction of a material directly to its wear rate. Detailed information on coefficients of friction for a wide range of materials is available in the article "Appendix: Static and Kinetic Friction Coefficients for Selected Materials" in this Volume.

The most prominent application of hard coatings for wear is in the life improvement of cutting tools. The CVD process for deposition of hard coatings was developed in the late 1940s and commercially applied to improve cutting tool life nearly 20 years later. One of the problems that had to be overcome to make CVD processing commercially feasible was

the need to reharden and temper the high-speed steel substrate after the CVD process was carried out at temperatures ranging from 1000 to 1100 °C (1830 to 2010 °F). This spurred the development of a low-temperature high-rate deposition process for hard coatings. Such a process (the ARE process) was developed by Bunshah and Raghuram in 1971 (Ref 16). The application of TiC and TiN coatings onto M42 high-speed steel tools by Bunshah and Shabaik (Ref 36) in 1975 was the first demonstration of large increases in cutting tool life. Bhusan (Ref 35) has compiled a table that summarizes tool wear-life improvement. Table 4 is an adaptation of Bhusan's original table.

Table 4 Improvement in cutting tool wear life obtained by deposition of a hard coating on the tool substrate

Tool substrate	Coating material/ treatment	Deposition method[b]	Coating thickness		Improvement in wear life[b]
			μm	**μin.**	
M-42 high-speed steel	TiC	ARE	5-8	200-320	3-8×
Cemented carbide	TiN, TiC	IP, CVD	5	200	IP Comparable to CVD
M10 high-speed steel[a]	TiC, TiN	ARE	2	80	20×
Cemented carbide	TiN	MS	5	200	Several times
Cemented carbide	TiC, TiN	CVD	8-10	320-400	Several Times
Cemented carbide	HfN, TiC/Al_2O_3, Al_2O_3, TiC/TiN	CVD	. . .	. . .	HfN superior
M7 high-speed steel[a]	TiN	MS	1-2	40-80	50×
Cemented carbide	(Ti, Al)N, TiN/TiC	CVD	. . .	. . .	3× better with (Ti, Al)N
M15 high-speed steel[a]	(Ti, Al)N, TiN	IP	. . .	. . .	3× better with (Ti, Al)N
Cemented carbide	TiN, HfN, ZrN	CVD, IP	8-15	320-600	Hardness of IP, MS, ARE; coatings at room temperature superior to CVD

(a) Twist drill.

(b) ARE, activated reactive evaporation; IP, conventional ion plating; CVD, chemical vapor deposition; MS, magnetron sputtering

Future Outlook

The most promising recent breakthrough in the field of multilayer coating technology appears to be the unbalanced magnetron sputter deposition process. This process should provide the next significant step in the improvement and evolution of state-of-the-art multilayer coating technology.

References

1. R.F. Bunshah, *Deposition Technologies for Films and Coatings,* R.F. Bunshah, Ed., Noyes Publications, 1982, p 85
2. S. Schiller, U. Heisig, and K. Goedicke, *Thin Solid Films,* Vol 40, 1977, p 327
3. R.G. Wilson and G.R. Brewer, *Ion Beams with Applications to Ion Implantation,* Wiley, 1973
4. C. Nishimura, K. Yanagisawa, A. Tago, and T. Toshima, *J. Vac. Sci. Technol.,* Vol A5 (No. 3), 1987, p 343
5. H.A. Beale and R.F. Bunshah, *Proceedings of 4th International Conference on Vacuum Metallurgy* (Tokyo), Iron and Steel Co. of Japan, p 238
6. R. Glang, *Handbook of Thin Film Technology,* L. Maissel and R. Glang, Ed., McGraw-Hill, 1970
7. L. Holland, *Vacuum Deposition of Thin Films,* Chapman and Hall, London, 1957
8. E.P. Graper, *J. Vac. Sci. Technol.,* Vol 10, 1973, p 100
9. R.F. Bunshah, *Deposition Technologies for Films and Coatings,* Noyes Publications, 1982, p 94
10. R.F. Bunshah, R. Nimmagadda, W. Dunford, B.A. Movchan, A.V. Dernchishin, and N.A. Chusanov, *Thin*

Solid Films, Vol 54, 1978, p 85
11. B. Rosa and L. Schultz, *Appl. Phys. Lett.,* Vol 53 (No. 16), 1978, p 1557
12. R.G. Humphreys, J.S. Satchell, N.G. Chew, and J.A. Edwards, *Appl. Phys. Lett.,* Vol 54 (No. 1), 1989, p 75
13. A.A. Karim, C.V. Deshpandey, H.J. Doerr, and R.F. Bunshah, *Thin Solid Films,* Vol 172, 1989, p 111-112
14. R.F. Bunshah, *Thin Solid Films,* Vol 107, 1983, p 21
15. R.F. Bunshah, U.S. patent 3,791,852, 1972
16. R.F. Bunshah and A.C. Raghuram, *J. Vac. Sci. Technol.,* Vol 9, 1972, p 1385
17. R.F. Bunshah and C.V. Deshpandey, *Physics of Thin Films,* J.L. Vossen and M.H. Francombe, Ed., Academic Press, 1987, p 60
18. D.M. Mattox, *Deposition Technologies for Films and Coatings,* R.F. Bunshah, Ed., Noyes Publications, 1982, p 244
19. M. Kobayashi and Y. Doi, *Thin Solid Films,* Vol 54, 1978, p 67
20. R.F. Bunshah, *New Trends in Material Processing,* American Society for Metals, 1978, p 200
21. A. Matthews and D.G. Teer, *Thin Solid Films,* Vol 80, 1981, p 41
22. Y. Murayama, *J. Vac. Sci. Technol.,* Vol 12, 1975, p 818
23. W. Kern and V.S. Ban, *Thin Film Processes,* J.L. Vossen and W. Kern, Ed., Academic Press, 1978, p 267
24. W. Kern and V.S. Ban, *Thin Film Processes,* J.L. Vossen and W. Kern, Ed., Academic Press, 1978, p 274
25. K.K. Yee, *Int. Met. Rev.,* Vol 1, 1978, p 19
26. H. Dun, P. Pan, F.R. White, and R.W. Douse, *J. Electrochem. Soc.,* Vol 128 (No. 7), 1981, p 1555
27. J.W. Coburn and M. Chen, *J. Appl. Phys.,* Vol 51, 1980, p 3134
28. J.C. Zesch, R.A. Lujan, and V.R. Deline, *J. Non-Cryst. Solids,* Vol 35-36, 1980, p 273
29. F. Jansen, *Plasma Deposited Thin Films,* J. Mort and F. Jansen, Ed., CRC Press, 1985, p 4
30. K. Yasui, H. Katoh, K. Komaki, and S. Kaneda, *Appl. Phys. Lett.,* Vol 56 (No. 10), 1988, p 898
31. R.W. Andreata, C.C. Abel, J.F. Osmundsen, J.G. Eden, D. Lubben, and J.E. Greene, *Appl. Phys. Lett.,* Vol 40 (No. 2), 1982, p 183
32. M. Minakata and Y. Furukawa, *J. Electron. Mater.,* Vol 15 (No. 3), 1986, p 159
33. J.W. Peters, F.L. Gebhart, and T.C. Hall, *Solid State Technol.,* Sept 1990, p 121
34. R.F. Bunshah, *IEEE Trans. Plasma Sci.,* Vol 18, 1990, p 846
35. B. Bhusan and B.V. Gupta, *Handbook of Tribology,* McGraw-Hill, 1991
36. R.F. Bunshah and A.H. Shabaik, *Research and Development,* Vol 26, 1975, p 46

Ion Implantation

George R. Fenske, Materials and Components Technology Division, Argonne National Laboratory

Introduction

ION IMPLANTATION is one of a number of surface modification processes that is emerging as an economical and viable process for improving the near-surface tribological properties of engineering components. In many engineering situations, material selection is often based on a compromise of bulk mechanical properties and near-surface tribological properties, with neither at their optimum values. As a result, there is considerable interest in fabrication processes, such as ion implantation, that make it possible to retain the bulk properties of a given compound yet still achieve desirable tribological properties in near-surface regions.

Ion implantation is a process by which virtually any element in the periodic chart can be injected into near-surface regions of any solid using a beam of high-velocity ions with energies typically ranging from 10 keV up to several MeV. As the

ions slow down in the material, they are distributed at depths ranging from a few nanometers to several micrometers, depending on the particle energy, angle of incidence, and substrate composition. Depending on the ion type, mass, energy, dose, deposition temperature, and substrate composition, the chemical, electrical, thermal, microstructural, and crystallographic properties of near-surface regions can be significantly altered to improve the friction and wear performance of the component.

Some of the advantages and limitations of ion implantation relative to other surface modification processes are listed in Table 1. For tribological applications, the more significant advantages include the low processing temperatures associated with ion implantation, allowing retention of critical bulk properties such as toughness and hardness. Another advantage is that the dimensions of the treated component are not changed significantly, thereby negating the need for subsequent refinishing steps on components with tight tolerances. Technically, the main disadvantage of ion implantation is that it is a line-of-sight process, making manipulation of components with complex shapes necessary to ensure uniform implantation of critical surfaces. However, it should be noted that recent advances in plasma-source technology have led to a novel ion-implantation process called plasma-source ion implantation that is not limited to line-of-sight implantations from one direction. This process will be discussed later in this article.

Table 1 Advantages and limitations of ion implantation

Advantages	Limitations
Low-temperature process	Shallow penetration depths
Retention of bulk properties	Relatively high capital and operating costs
No significant dimensional changes	Line-of-sight process
Solid solubility limit can be exceeded	Peak concentration limited by sputtering processes
Preparation of alloys independent of thermal diffusion	
Permits fast screening of alloy composition effects	
No sharp interfaces, reducing adhesion problems	
Controllable depth distribution of implanted species	
Clean vacuum process	

Ion implantation is a nonequilibrium process that generates a large number of mobile defects in near-surface regions in addition to deposition of the implanted species. In order to understand how these implanted ions and defects affect the tribological performance of near-surface regions, it is necessary to understand the effect of ion implantation on microstructure and mechanical and chemical properties, as illustrated in Fig. 1. The following sections describe the basic physical phenomena that occur during implantation of energetic ions, how these phenomena affect the properties of the implanted regions, and how these properties influence the tribological performance of near-surface regions. A description of the physical equipment used to implant components, critical implantation parameters, and examples of current commercial applications are also provided.

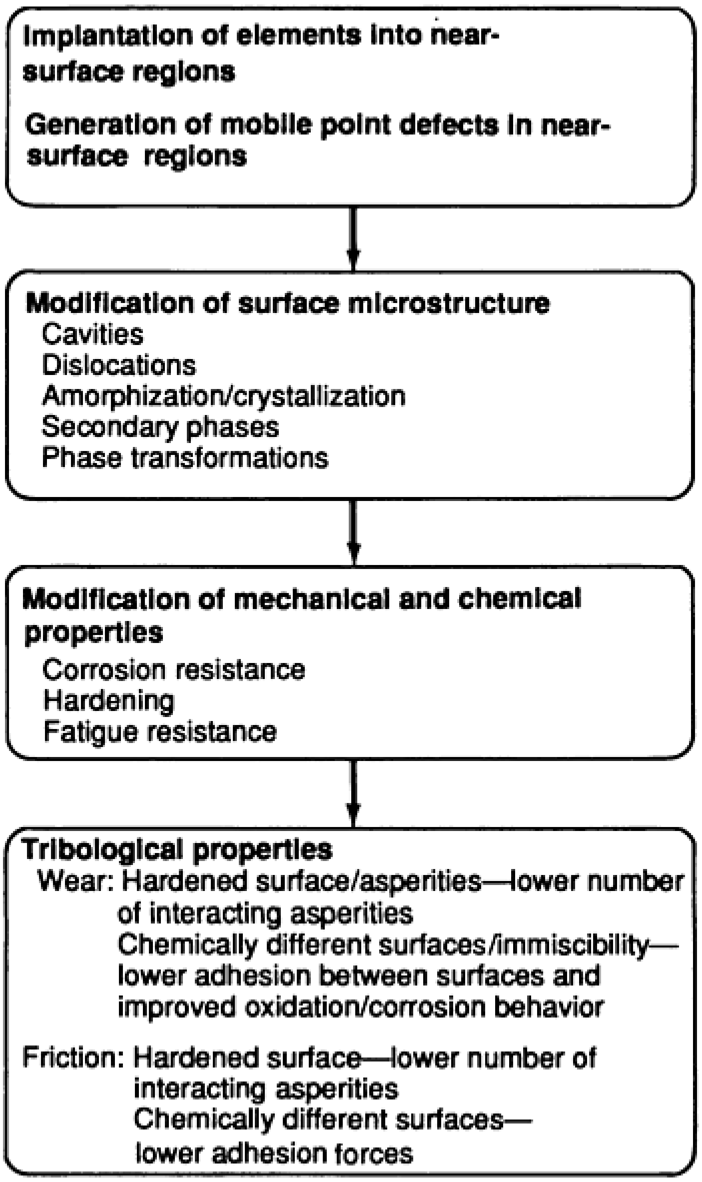


Fig. 1 Flow diagram illustrating mechanisms by which defects produced during ion implantation lead to improved tribological properties

Acknowledgement

This work was supported by the Tribology Program, Office of Transportation Materials, U.S. Department of Energy, under Contract No. W-31-109-ENG-38.

Ion Implantation Processes

Basic Principles. Ion implantation involves creating a charged particle (for example, C^+ or $Ti^{+,2+,3+}$) in an ion source, accelerating the ion to high velocities, and directing a beam of ions onto the substrate, where the ions dissipate their energy and come to rest by colliding with substrate atoms. The resulting defects produced by these collisions, together with the presence of implanted species, interact via a number of processes to produce microstructural changes in the near-surface region that result in altered mechanical and chemical properties.

Figure 2 illustrates phenomena that occur as the ion impinges on the near-surface region of the substrate. As the ions penetrate into the substrate, they undergo numerous collisions with atomic nuclei and electrons, during which energy is transferred to the atoms and electrons. Collisions with atoms in the substrate occur at discrete locations, changing the

direction of the incident ion and resulting in a loss of energy. If the amount of energy transferred to a substrate atom is sufficiently large, the substrate atom is displaced from its original lattice site and is typically called a primary knock-on atom (PKA). Depending on the energy transferred during a collision, a PKA can itself cause displacements of lattice atoms (termed secondary knock-on atoms, or SKAs).

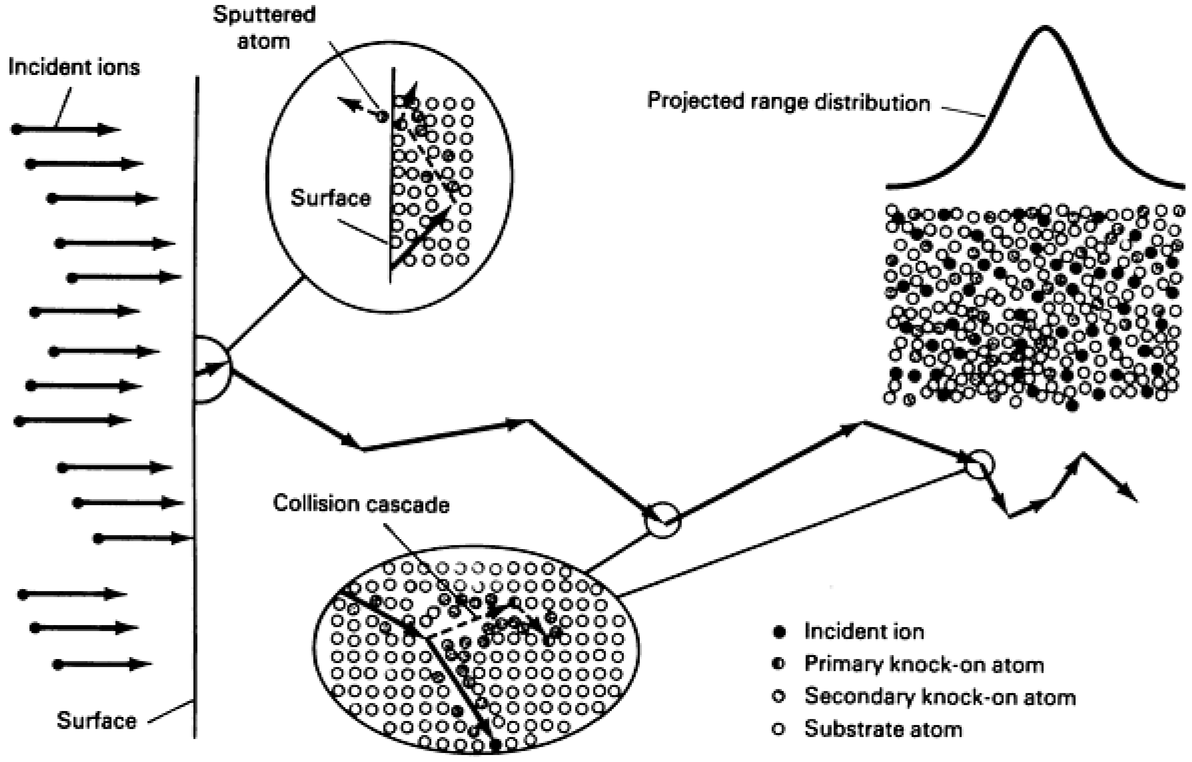


Fig. 2 Schematic of phenomena that occur in near-surface regions during ion implantation

The production of PKAs and SKAs continues until the energy of the respective atoms is below a critical energy where the moving atom/ion can no longer displace another lattice atom (typically on the order of 20 to 50 eV). Because of the scattering kinetics, SKAs are typically produced in a very localized region, called a cascade, around the site where the PKAs are produced. The path of the incident ion will therefore consist of a jagged line along which are distributed a number of cascades of displaced substrate atoms (Fig. 2).

The energy of the incident ion is transferred to both atoms (nuclei) and electrons. The energy transferred to the atoms can be thought of as occurring in the discrete collisions between the incident ion and the PKAs and SKAs. In collisions with electrons, the mass of the electrons is so small compared with that of the moving ions (or PKAs and SKAs) that kinematically very little energy is lost during a single collision. The small mass of the electron also means that the direction of the moving atom is not altered by the collision.

Range and Energy Profiles. Two important parameters that provide valuable insight on the depth of the modified region are the projected range distribution of the implanted ion and the energy deposition profile. The projected range distribution is the probability (per unit depth increment), $P(x)$, that an incident ion will come to rest in the substrate at a distance, x, from the surface of the substrate. The projected range distribution is a function of the atomic number, mass, energy, and angle of incidence of the incident ion and the atomic numbers, masses, and densities of the elements in the substrate. The distribution profiles are usually quoted for ions that have come to rest in the substrate; however, it is possible to calculate distribution functions of ions as they slow down in the substrate--that is, the distribution function of an ion that began with an incident energy, E, and has slowed down to an energy E'. The concentration profile, or atomic density, of the implanted ions, $N(x)$, is the product of the projected range profile and the ion dose, ϕ:

$$N(x) = P(x)\,\phi \qquad \textbf{(Eq 1)}$$

where $N(x)$ is in ions/cm^3, $P(x)$ is in cm^{-1}, and ϕ is in ions/cm^2.

The energy deposition profile represents the energy deposited in the substrate as a function of depth. The energy can represent the energy associated with displacing lattice atoms or the energy associated with ionization processes. A useful parameter that can be derived from the energy deposition (energy associated with displacing atoms) profile is the atomic displacement profile, defined as the average number of times each lattice atom is displaced. Kinchin-Pease (Ref 1) used a simple model to predict the number of atoms displaced by a PKA of energy E as:

$$\nu(E) = \frac{CE}{2E_d} \qquad \textbf{(Eq 2)}$$

where $v(E)$ is the number of displaced lattice atoms due to a PKA of energy E, C is a constant (typically 0.8), and E_d is the energy required to displace a lattice atom (typically 20 to 30 eV). The displacement profile, $G(x)$, is thus the product of the energy deposition function and $C/(2E_d)$:

$$G(x) = \frac{F(x)\,\phi\,C}{2E_d} \qquad \textbf{(Eq 3)}$$

where $G(x)$ is the number of displaced lattice atoms (per unit volume), $F(x)$ is the energy consumed in displacing atoms (eV per unit distance), ϕ is the incident ion flux (number per unit area), and E_d is the displacement energy (eV). Dividing $G(x)$ by the atomic number density (atoms per unit volume) gives the average number of times a lattice atom is displaced, or the displacements per incident atom (commonly called dpa):

$$\mathrm{dpa}(x) = \frac{G(x)}{N} \qquad \textbf{(Eq 4)}$$

where dpa(x) is the average number of times a lattice atom is displaced as a function of depth from the surface and N is the atomic number density of the lattice atoms.

Range and energy (or dpa) profiles can be calculated by analytical treatments or by Monte Carlo simulations of ion-scattering processes. Analytical treatments use spatial moments of the profiles derived from transport theory to reconstruct the depth distributions (Ref 2, 3, 4, 5, 6, 7, 8, 9, 10). Monte Carlo simulations follow the trajectories of a large number of ions as they penetrate and slow down in the lattice (Ref 11, 12). Depth profiles are obtained from Monte Carlo calculations by keeping track of the number of events (for example, lattice atom displacements) that occur in a given depth interval for a large number of incident ions.

Figures 3(a) and (b) show examples of the first and second spatial moments (average projected range and straggling) as functions of the incident-ion energy for nitrogen and titanium ions, respectively, incident on an iron substrate (Ref 11). Note that the range increases with increasing energy; that is, the greater the incident energy, the greater the penetration of the ion into the substrate. Also note that the nitrogen ion penetrates farther than the titanium ion for a given energy (lighter ions penetrate farther than heavy ions). The depth of penetration and hence the approximate depth of the affected zone for 100 keV ions, which are typically used in many tribological implantations, is approximately 40 to 50 nm for titanium and approximately 100 nm for nitrogen in iron.

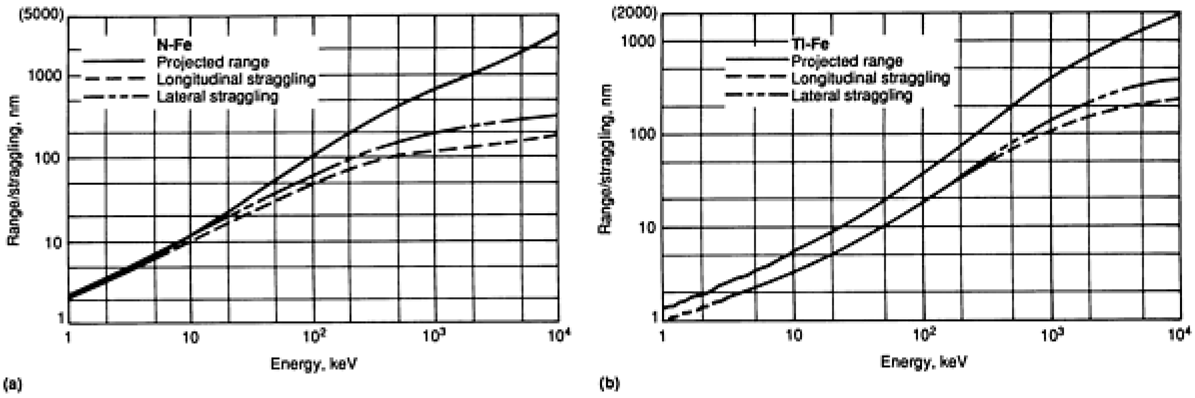


Fig. 3 Examples of projected range and straggling as functions of incident-ion energy. (a) Nitrogen ions incident normal to an iron substrate. (b) Titanium ions incident normal to an iron target. Source: Ref 11

Figures 4(a) and (b) show projected range profiles (solid curves) of 100 keV nitrogen and titanium ions implanted into iron. The profiles were constructed using a Pearson IV distribution function (Ref 13). Also shown are Monte Carlo calculations (based on the trajectories of 10,000 incident ions) of the projected range profiles (denoted by the histograms) and the atomic displacement profiles (dashed curves) (Ref 11). The atomic displacement profiles in Fig. 4 clearly illustrate that 100 keV titanium ions displace considerably more lattice atoms than 100 keV nitrogen ions.

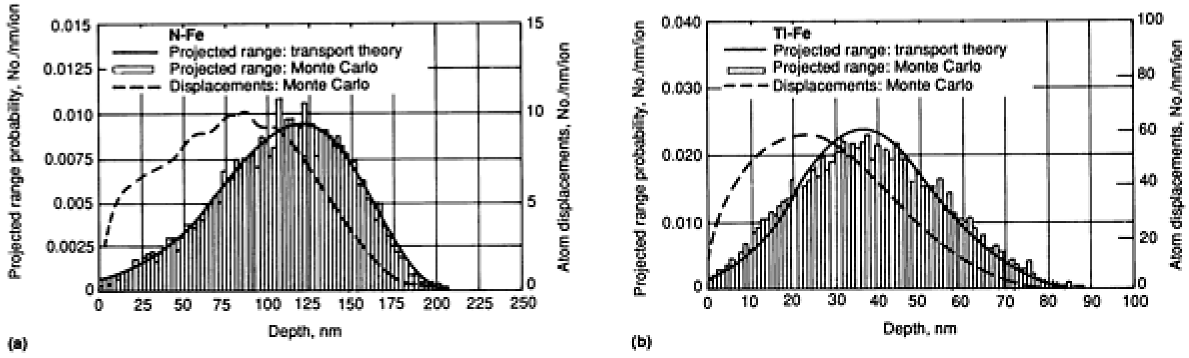


Fig. 4 Damage and projected range distributions of 100 keV nitrogen ions (a) and 100 keV titanium ions (b) incident normal to an iron substrate. Source: Ref 11, 13

Sputtering. Another phenomenon that occurs at the surface, called sputtering, involves the gradual erosion or removal of material from the top surface as it is bombarded by energetic particles. When a collision occurs between an incident ion and a lattice atom or between a PKA and a lattice atom at or near the surface, sufficient momentum can be imparted such that the lattice atom is ejected from the surface. This eventually leads to a gradual erosion or removal of material (both implanted and lattice atoms) from the surface. Subsequent ions are therefore implanted deeper into the surface (measured from the original location of the surface). This has a net effect of limiting the maximum atomic concentration of the ions that can be implanted in near-surface regions.

Figures 5(a) and (b) illustrate the effect of sputtering on the final implant profile for 100 keV nitrogen and titanium implantations, respectively, into an iron substrate (Ref 13). An important parameter in calculating the effect of sputtering on the final implant profile is the rate at which lattice atoms are removed by sputtering. This rate, or sputtering yield, is defined as the number of lattice atoms ejected from the surface per incident ion. The sputtering yield can be calculated analytically (Ref 14), semiempirically (Ref 15), or with Monte Carlo codes (Ref 11), such as those used to generate the projected range and atomic displacement profiles in Fig. 4. The sputtering yield is dependent on the incident ion, energy, substrate composition, and angle of incidence. For the calculations in Fig. 5, the sputtering yields of 100 keV nitrogen and titanium ions incident normal to an iron substrate were calculated to be 0.42 and 4.7 atoms/ion, respectively (Ref 13).

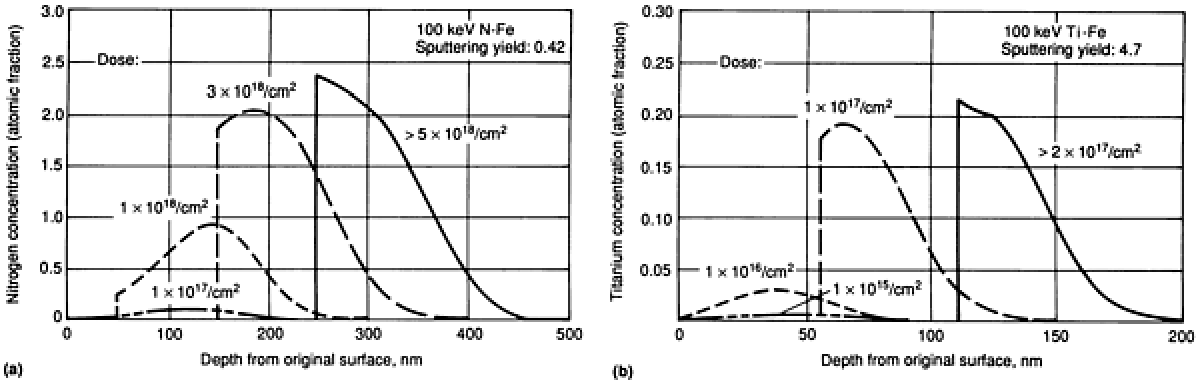


Fig. 5 Concentration profiles of 100 keV nitrogen ions (a) and 100 keV titanium ions (b) implanted into an iron substrate at high doses. Source: Ref 13

At a dose of 1×10^{17} nitrogen ions/cm^2 (Fig. 5a) and a yield of 0.42, approximately 5 nm of material will be removed because of sputtering. As a result, the implant profile is very similar in shape to the projected range profile shown in Fig. 4(a). As the dose is increased, more and more material is sputtered, producing drastic changes in the implant profiles, as shown for the high-dose implantation profiles in Fig. 5(a). At a dose of 3×10^{18} nitrogen ions/cm^2, approximately 75 nm has been sputtered from the surface, and the peak nitrogen concentration of approximately 200 at.% is still located below the surface. However, for doses above 5×10^{18} ions/cm^2, the peak nitrogen concentration of approximately 230 at.% lies at the surface of the sputtered substrate. Any further increase in dose beyond 5×10^{18} nitrogen ions/cm^2 will in essence be wasted effort, because sputtering will continue to remove previously implanted nitrogen ions as quickly as they are implanted.

A similar effect is observed for the titanium ion implantations shown in Fig. 5(b), except that the doses are decreased since the sputtering yield is 4.7 for 100 keV titanium ions incident on iron compared with 0.42 for 100 keV nitrogen ions. As shown by Fig. 5(b), doses in excess of 2×10^{17} titanium ions/cm^2 provide a concentration profile in which the peak concentration lies at the surface. It can also be seen that the peak titanium concentration at 2×10^{17}/cm^2 is approximately 20 at.% versus approximately 230 at.% for the high-dose nitrogen implants. Thus, the major effects of sputtering are to shift the location of the peak concentration closer to the exposed surface and to limit the maximum amount of material that can be implanted into near-surface regions.

Microstructural Properties of Implanted Regions

Near-Surface Region Defects. Implantation of energetic ions into components produces a large number of defects, such as vacancies, interstitials, and the implanted species, in the near-surface region that, depending on the deposition temperature, are potentially highly mobile even after they are in thermal equilibrium with the surrounding matrix. Although most of the vacancies and interstitials (and small clusters of vacancies and interstitials) generated in cascades diffuse through the matrix and recombine (for example, a vacancy and an interstitial may react to form the equivalent of a normal lattice site), sufficient numbers avoid recombination to alter the microstructure of the implanted region.

The mobile defects that survive react with existing microstructural features, such as dislocations, grain boundaries, free surfaces, secondary phases, and voids/bubbles, or combine with other defects to nucleate such microstructural features as dislocation loops, voids, and so forth. For example, Fig. 6 shows cross-sectional transmission electron microscopy (XTEM) micrographs of the near-surface region of nickel foils that were implanted at room temperature with energetic (500 keV) helium ions at two different doses: 2×10^{15}ions/cm^2 (Fig. 6a) and 1×10^{17}ions/cm^2 (Fig. 6b). At the lower dose, small dislocation loops can be observed in a region corresponding to a depth where the damage profile is at a maximum. At the higher dose, the dislocations extend to the surface of the foil (denoted as the interface in Fig. 6b).

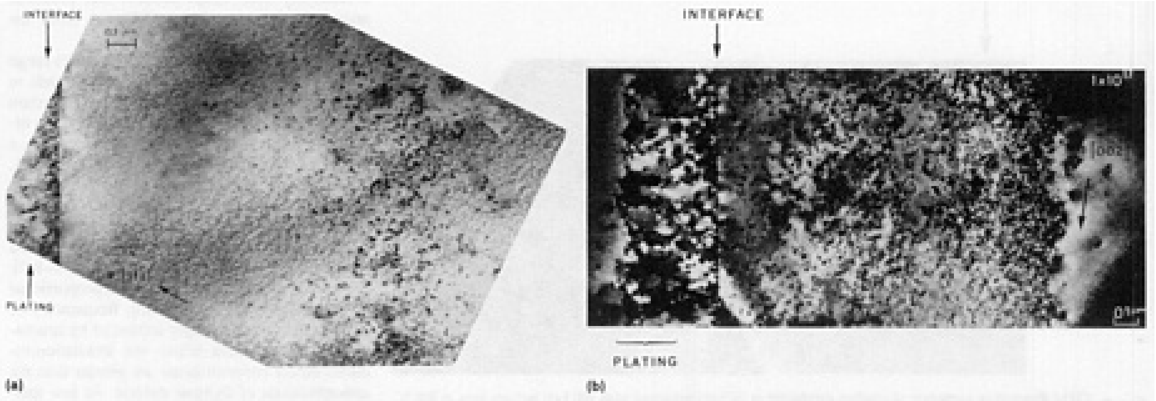


Fig. 6 XTEMs illustrating dislocations in near-surface regions of nickel implanted with 500 keV helium ions. (a) 2 × 10^{15}ions/cm^2. (b) 1 × 10^{17}ions/cm^2. Source: Ref 16

At higher temperatures and higher doses (500 °C, or 930 °F, and 5 × 10^{17}ions/cm^2, as shown in Fig. 7), the density of dislocations increases compared with those in Fig. 6. Typically, the dislocation density will increase at higher doses and decrease at higher temperatures, assuming other deposition conditions remain constant. In Fig. 7, the increase in the dose had a greater effect on the dislocation density than the increase in temperature.

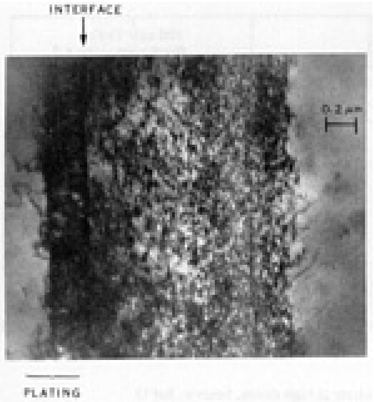


Fig. 7 XTEM illustrating high density of dislocations in a nickel foil implanted with 500 keV helium ions at 500 °C (930 °F) to a dose of 5 × 10^{17}ions/cm^2. Source: Ref 16

Under proper conditions, small three-dimensional voids and bubbles nucleate and grow in the irradiated region, as shown in Fig. 8. In this case (500 °C, or 930 °F, and 1 × 10^{17}ions/cm^2), the cavities range from 3 to 10 nm in size, depending on the depth, and occupy up to 1% of the volume. As higher doses are administered, the cavities grow and coalesce to form larger cavities. Eventually, the pressure inside the cavities, which are filled with the implanted helium ions, becomes sufficiently large to plastically deform the matrix above the cavities and form blisters* on the top surface.

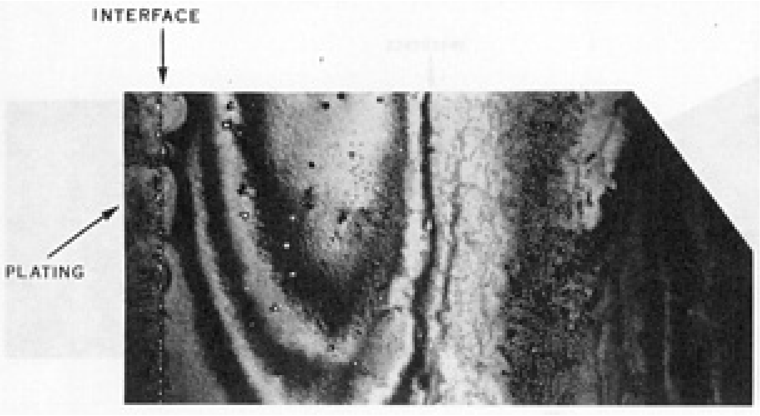


Fig. 8 XTEM illustrating presence of cavities produced in nickel implanted with 500 keV helium ions at 500 °C (930 °F) to a dose of 1 × 10^{17}/cm^2. Source: Ref 16

Crystalline-to-Amorphous Transformation. Ion implantation of energetic particles can sometimes change the crystalline nature of near-surface regions, producing amorphous layers (Ref 17, 18, 19, 20, 21, 22, 23, 24, 25) or, alternatively, can induce crystallization in near-surface regions of amorphous compounds (Ref 25, 26, 27). Several models have been proposed to account for the transformation from a crystalline to an amorphous state. One model suggests that amorphization is a result of rapid quenching (on the order of 10^{14} K/s) on thermal spikes produced in collision cascades (Ref 22). However, high-energy electron irradiations--which produce individual Frenkel-pair defects, but no cascades--are also capable of producing amorphous regions. Another model (applied to amorphization of ordered intermetallic alloys) treats amorphization as thermodynamically similar to melting phenomena and predicts that its onset occurs at a critical volume expansion, suggesting that amorphization is caused by a mechanical instability (Ref 17). The onset of amorphization and the degree to which it occurs depends on a number of factors, including temperature, implant species and substrate composition, dose, and dose rate.

Secondary Phase Formation. Under appropriate conditions, ion implantation can also lead to the formation of secondary phases that improve the tribological properties of the near-surface region. For example, nitride, carbide, oxide, and boride precipitates have been observed in both pure iron and steels after high-dose implantation with nitrogen, carbon, oxygen, and boron, respectively (Ref 28, 29, 30, 31, 32, 33, 34). In some instances, a postimplantation annealing treatment is employed; examples of such cases include the implantation of metallic ions such as molybdenum, iron, and nickel into metallic alloys, leading to the formation of intermetallic phases (Ref 35, 36, 37, 38), and the implantation of nitrogen into titanium alloys (Ref 39).

Modeling of Microstructural Changes. Research in the modeling of microstructural changes during irradiation has been quite extensive (Ref 39, 40, 41, 42). The overall approach utilizes the approximation of (quasi-) chemical or reaction-rate theory to calculate time- and spatially dependent concentrations of defects (for example, vacancies, interstitials, vacancy clusters, and dislocations), resulting in a set of equations of the type (Ref 43):

$$\frac{dC_j}{dt} = \Sigma_j\, w(k;j)C_k - \Sigma_k\, w(j;k)C_j + G_j - L_j \quad \text{(Eq 5)}$$

where C_j is the concentration of defects or clusters of type j (for example, voids containing j vacancies or dislocation loops containing j interstitials), $w(k;j)$ is the rate per unit concentration of transitions of defects or clusters of type k to type j, G_j is the rate at which defects or clusters of type j are formed directly (for example, formation of a dislocation loop directly from a collision cascade), and L_j is the loss rate of defects of size j by diffusion to sinks (for example, incorporation of vacancy clusters into large cavities). Calculation of the reaction parameters, $w(j;k)$, requires detailed knowledge of the physics of the different mechanisms involved in the evolution of the microstructure--for example, diffusion of the different species, interaction of different species, characteristics of defect generation, and interactions with defect sinks.

Solutions to these very approximate equations are straightforward (with many simplifying assumptions), but extremely time consuming because of the large number of coupled equations that must be solved for each type (and size *j*) of defect cluster (vacancies, interstitials, and implanted species). In compounds irradiated with energetic ions, the approach typically involves solving sets of coupled differential equations of the general form (Ref 43, 44):

$$\frac{\partial C_j(x,t)}{\partial t} = \frac{D_j \partial^2 C_j(x,t)}{\partial x^2} + S_j(x,t) \quad \textbf{(Eq 6)}$$

where C_j is the time- and depth-dependent concentration of elements (or defects) of type *j,* and S_j represents local rates of production and loss of species of type *j* by various processes, such as direct production of vacancies or interstitials during the slowing-down process of implanted ions, trapping of solute atoms at sinks, incorporation of defects into larger defects, and formation or dissolution of precipitates.

Solutions to the sets of differential equations (Eq 6) necessary to determine compositional changes in near-surface regions require modeling a number of physical phenomena (Ref 42, 45). These phenomena include (in addition to the chemical compositional change caused by addition of the implanted species) athermal processes, such as preferential sputtering and displacement mixing, and thermally activated processes, including radiation-enhanced diffusion, radiation-induced segregation, and Gibbsian adsorption.

Preferential sputtering is caused by differences in the sputtering yields of the different elements that comprise an alloy. Because of differences in the momentum transferred to elements of differing masses, the energy available to overcome a critical binding energy will differ from element to element in an alloy. In addition, differences in the chemical interactions of the different constituents in the alloy will result in different critical binding energies from element to element. As a result of these two effects, the individual sputtering yields of the different elements that constitute the material can differ, leading to an enrichment of elements with low yields and a depletion of elements with high yields. Because sputtered atoms typically originate from the first one or two monolayers, this is a relatively short-range process that produces near-surface compositions different from the bulk composition.

Gibbsian Adsorption. Another short-range process is Gibbsian adsorption, which tends to increase or decrease the concentration of certain alloying elements in the top surface layer in order to minimize the surface free energy. In a binary alloy consisting of elements A and B, Gibbsian adsorption would result in the top monolayer being enriched in either element A or B relative to the bulk composition. Because Gibbsian adsorption is dependent on the forward bulk-to-surface and backward surface-to-bulk jump frequencies, which in turn are proportional to the concentrations and jump frequencies of point defects, it is strongly enhanced by irradiation at temperatures where the irradiation-induced defect concentrations are greater than the concentrations of thermal defects. At low temperatures, Gibbsian adsorption is also affected by displacement-mixing-induced motion of atoms within cascades produced at the surface.

Gibbsian adsorption and preferential sputtering processes tend to counteract each other (Ref 45). For example, if in a binary alloy, element A is preferentially sputtered relative to element B, while Gibbsian adsorption produces a surface layer enriched (relative to the bulk composition) in element A, the two effects will counteract each other. However, if element B is preferentially sputtered while A is enhanced by Gibbsian adsorption, both effects will still counteract each other, because the increased concentration of A on the surface caused by Gibbsian adsorption temporarily produces a higher flux of A atoms from the surface, resulting in a pseudopreferential sputtering yield of A that counteracts the preferential yield of B.

Displacement mixing is an athermal process that results in a redistribution of alloying elements because of energetic displacement of atoms in collision cascades. The process can be modeled in a manner similar to diffusion, with the diffusion coefficient being dependent on the spatially dependent damage (or dpa) rate, the nearest-neighbor distance, and the number of atoms changing sites per dpa (Ref 46).

Radiation-Enhanced Diffusion. At elevated temperatures, thermally activated diffusional processes become dominant. The diffusional processes are enhanced by the excess concentrations of mobile vacancies and interstitials that survive recombination processes in cascades and at internal sinks. This process, termed radiation-enhanced diffusion, typically is dominant at temperatures between 0.2 and $0.6T_m$, where T_m is the melting point. Above approximately 0.6 T_m, the thermodynamic equilibrium concentration of vacancies becomes larger than the radiation-produced concentration of defects, and thus the diffusion coefficient is determined by conventional thermally produced vacancies. Below

approximately $0.2T_m$, the concentration of radiation-produced vacancies may be high, but the vacancies are not as mobile and the diffusion process is dominated by displacement mixing.

Radiation-Induced Segregation. In contrast to radiation-enhanced diffusion processes, which tend to flatten out composition gradients, radiation-induced segregation processes tend to increase local composition gradients (Ref 21, 44, 47). When the local concentration is increased above local solubility limits, radiation-enhanced segregation can lead to formation of phases normally not observed for the bulk alloy composition. Radiation-induced segregation is caused by a preferential coupling of alloying elements with defects to form mobile defect-solute complexes. Defect fluxes that arise from gradients in defect concentrations during irradiation thus induce net fluxes. Gradients in the defect concentrations that arise from spatially nonuniform production of defects, as well as gradients induced by the presence of sinks (the top surface, grain boundaries, or defect clusters such as dislocations, precipitates, or cavities), produce persistent fluxes of defects and thus a net flow of defect-solute complexes.

Radiation-induced segregation is strongly temperature dependent (Ref 48); at low temperatures, the high concentration of defects coupled with their low mobility increases their recombination, thus subduing their long-range migration and segregation. On the other hand, normal diffusion processes at high temperatures tend to flatten out compositional gradients, thereby also subduing segregation. Thus, radiation-induced segregation is dominant only at intermediate temperatures, where strong coupling of solute atoms with mobile defects exists.

An example of radiation-induced segregation is illustrated in Fig. 9, which shows the formation of (γ')Ni_3Si (25 at .% Si) from a nickel-silicon alloy with a bulk silicon concentration of 6 at .%. The top surface is the dominant sink for defects produced in near-surface regions, setting up a net flux of defects toward the surface. Because silicon in nickel-silicon alloys is preferentially transported in the same direction as the defect fluxes (Ref 21), silicon is transported to the surface of the implanted film. When the concentration of silicon exceeds the solubility limit (approximately 10 at .% in nickel-silicon), a coating of Ni_3Si forms on the top surface. Farther from the surface, grain boundaries and dislocation loops act as internal sinks for defects, producing net fluxes of defects (and Si) to them. Thus, the formation of Ni_3Si coatings on these sinks is also observed.

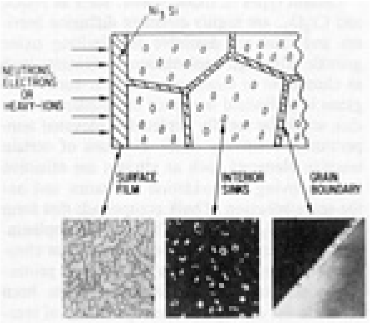


Fig. 9 Radiation-induced segregation in a Ni-6Si(at.%) alloy implanted with MeV nickel ions at elevated temperatures. Source: Ref 21

The important aspect of this discussion is that the microstructure of the implanted region can be significantly different from the starting microstructure. This can be attributed in part to the direct implantation of a foreign atom into the surface, thereby altering its chemical composition and microstructure. However, the microstructure can also be altered dramatically by the presence of mobile defects that interact by a number of mechanisms. This is readily seen in the alloy shown in Fig. 9, which was implanted not with foreign atoms, but with nickel atoms--the same element that constituted 94 at .% of the bulk composition. It is also important to note that the damage rate (or current density on the substrate) and temperature play major roles in determining the rates at which these processes occur.

Note cited in this section

* * Under most conditions encountered in the use of ion implantation in tribological applications, blister formation is not of concern, because the implanted species are either soluble in the matrix, are not gaseous, or, if they are gaseous (for example, implantation of nitrogen into titanium alloys), form solid compounds (for example, TiN).

Effect of Microstructure on Tribological Properties

The microstructural and chemical changes caused by ion implantation can significantly affect the tribological performance of engineering components. The extent of this influence depends on the type of wear/friction mechanisms present--for example, adhesive, abrasive, fatigue, or chemical interactions (Ref 49, 50, 51, 52, 53).

Adhesive and Abrasive Wear. For adhesive or abrasive wear conditions, it is important to note that when two macroscopically smooth surfaces are brought together, contact is made at isolated asperities on the mating surfaces. Models for adhesive or abrasive wear assume that the wear volume (or wear rate) is proportional to the number of asperities in contact (Ref 49), which in turn depends on the applied load and the yield strength of the asperities. The higher the yield strength, the fewer the number of asperities required to support the applied load. Thus, a lesser amount of material will be removed because of the lower number of asperities in contact. As a result, the wear volume is often found to be inversely proportional to the hardness (where the hardness is proportional to the yield strength) of the materials in contact. One mechanism by which ion implantation improves the wear behavior is strengthening of near-surface regions, which reduces the number of asperities that come into contact.

Strengthening Mechanisms. A number of mechanisms based on immobilizing dislocations have been examined in terms of their ability to strengthen ion-implanted surfaces (Ref 50). These mechanisms include cold working, substitutional solid-solution strengthening, interstitial solid-solution strengthening, second-phase strengthening, and phase transformations.

Conventional cold working of metals is often used to create a high density of tangled dislocations that restrict the motion of mobile dislocations by introducing strain fields and dislocation locks, thereby increasing the yield strength. As seen in Fig. 7, ion implantation can produce a high density of tangled dislocations. In addition, at lower temperatures, the implanted material will also have a high density of frozen-in point defects that can restrict the movement of dislocations. Both of these effects can lead to stronger near-surface regions with improved friction and wear performance.

Substitutional solid-solution strengthening is achieved by forcing impurity atoms with large atomic size differences with respect to the host atoms into substitutional lattice positions, resulting in large strain fields that the moving dislocations must overcome. The yield strength increases with the concentration of the solute and the atomic size difference. Because of its ability to implant virtually any element in the periodic chart, ion implantation is ideally suited to introducing misfitting atoms into the host matrix and thus strengthening the surface region. Ion implantation may even have an advantage in substitutional solid-solution strengthening, because it can introduce large concentrations of misfitting atoms in excess of those thermodynamically possible.

Interstitial solid-solution strengthening is another mechanism by which ion implantation can be used to strengthen near-surface regions and thus improve tribological performance. It is analogous to substitutional solid-solution strengthening; in this case, however, the interstitials are often mobile and diffuse to dislocations where they may become trapped and thus raise the energy barrier for dislocation motion. Carbon and nitrogen are two commonly implanted elements that act as interstitials in a number of alloys. It should be noted that a postimplantation, low-temperature aging treatment may be required for this effect to be observed (Ref 50).

Second-phase strengthening is another mechanism by which ion implantation can harden near-surface regions. In this case, the formation of fine precipitates (for example, by implantation of carbon into titanium alloys to form TiC precipitates) provide a barrier for dislocation motion and thus strengthen the surface.

Finally, phase transformation of a near-surface region to an entirely different phase by implantation of appropriate ions (for example, high-dose nitrogen implantations into chromium to produce CrN) can also improve tribological properties.

In this case, the improvement is not necessarily caused by limiting dislocation motion in the parent material, but rather by transforming a large portion of the implanted region into a different, stronger phase.

All of these mechanisms improve the wear behavior of implanted compounds by strengthening near-surface regions and thereby decreasing the number of asperities required to support the applied load. This in turn reduces the number of surface asperities involved in the wear process and hence the volume of material removed. This effect is important in both adhesive and abrasive wear environments. For adhesive wear, which is caused by local welding of asperities that come into contact, the ability of the asperities to adhere to one another is also critical. It has been suggested that surfaces composed of elements that are readily soluble in one another have high adhesion to each other and thus should have higher adhesive wear rates than two surfaces composed of elements that are insoluble in one another (Ref 51). For example, silver has limited solid solubility (<0.1%) in iron, while titanium is soluble (>1%) in iron; thus, a silver asperity should not weld itself to a mating iron asperity as readily as a titanium asperity. Ignoring other considerations (such as differences in the hardnesses of the asperities), it follows that greater adhesive wear losses can be anticipated for the titanium-iron couple than for the silver-iron couple. The importance of this observation is that ion implantation can also modify the chemical composition of near-surface regions (and hence surface asperities) in such a manner as to reduce the mutual solubility and adhesion of mating asperities.

Fatigue wear involves the initiation and propagation of cracks in materials by the application of cyclic stresses. The crack initiation stage is sensitive to the condition of the surface and thus can be modified by ion implantation. The crack propagation stage extends deep into the bulk of the substrate relative to the depth of the implanted ions and thus is normally not affected by ion implantation. It is during the crack initiation/nucleation stage, therefore, where ion implantation has a role in reducing fatigue by limiting dislocation movement in near-surface regions, thereby preventing persistent slip bands from forming a surface topography conducive to crack nucleation. The mechanisms by which dislocation motion and crack nucleation are reduced are similar to those described for strengthening near-surface regions. Thus, the methods for improving adhesive and abrasive wear properties are also applicable for improving fatigue life.

Corrosive wear processes involve the formation of corrosion products that remove surface material by mechanical action or by spallation arising from excessive internal stresses during film growth. Oxidation is one type of corrosion that can have a significant impact on the wear of components, particularly those fabricated from steels. In many instances, some degree of corrosion is desirable if the corrosion film prevents strong adhesion between mating asperities and if the corrosion film is hard and tenacious. However, if the corrosion film has high growth rates, or is brittle and does not adhere, the surface will undergo excessive wear.

Mechanisms Affecting Oxidation Resistance. Four mechanisms have been identified by which alloying (and hence ion implantation) can modify oxide growth rates (Ref 24): modification of the oxide defect structure, modification of the protective layer by incorporation of reactive elements, blocking mechanisms, and modification of the oxide plasticity. Modification of the oxide defect structure by ion implantation involves the incorporation of donor impurities that alter the electronic conductivity of the oxide, thereby modifying the electrochemical driving forces for diffusion of cations, anions, and electrons and hence growth of the oxide layer.

Certain types of oxide scales, such as Al_2O_3 and Cr_2O_3, are highly effective diffusion barriers and thus are desirable for limiting oxide growth. Although implantation of elements such as chromium or aluminum into near-surface regions is of limited use since the implanted species would be rapidly depleted at elevated temperatures, relatively small additions of certain reactive elements such as yttrium are effective for improving the oxidation resistance and oxide-scale adhesion of bulk compounds that form Al_2O_3 or Cr_2O_3 scales. Thus, while implantation of the large quantities of aluminum or chromium necessary to form a long-lasting protective scale may not be feasible, it has been effective for adding the small quantities of reactive elements required to enhance the adhesion and endurance of Al_2O_3 and Cr_2O_3 films.

Grain boundaries and dislocations act as paths for rapid diffusion of oxygen ions during oxidation. Ion implantation is effective in limiting diffusion along these paths by incorporating oversized solute atoms that are trapped at these fast-diffusion paths, thereby blocking the motion of oxygen ions.

The ability of an oxide film to relieve growth stresses will depend on its plasticity, which in turn depends on dislocation climb. Implantation of smaller cations that have a higher diffusivity than the host cation in an oxide lattice should therefore improve the ability of the oxide film to plastically deform, thus relieving growth stresses and thereby improving its adhesion to the underlying substrates. However, at higher doping levels, or for cations too small to be accommodated into the lattice, a second-phase oxide may form and pin dislocations, leading to lowered plasticity. It is also important to note that because plasticity is lower in stoichiometric oxides, doping in order to reduce defect concentrations (and hence diffusion) may actually reduce oxide plasticity.

Mechanisms Affecting Aqueous Corrosion. Implantation of atoms into near-surface regions also affects aqueous corrosion of metals by expanding the passive range of the surface, enhancing the resistance of the surface to localized corrosion, forming amorphous surface layers (thereby eliminating rapid corrosion at grain boundaries), shifting the open-circuit potential into a passive range where corrosion current densities are low, modifying the galvanic action of second-phase particles, and inhibiting the kinetics of cathodic reactions on the surface.

Although ion implantation can significantly affect the corrosive behavior of surfaces, its impact on corrosive wear is more complicated because of the mechanical stresses imposed on the contacting surfaces and because of frictional heating of asperities on a localized basis. Both of these factors affect the adhesion, ductility, and thermal expansion of the corrosion film. They are also important because of the cyclic nature of the wear process, particularly for sliding wear applications, which is affected by the growth rate of the corrosion film (that is, the amount of film growth that occurs between successive mechanical encounters of asperities).

Ion Implantation Equipment

Ion implanters consist of a number of components that generate, accelerate, and focus the ions onto a surface (Ref 54, 55, 56). Figure 10 illustrates several types of implantation systems, including mass-analyzed ion implanters (Fig. 10a), broad-beam implanters (Fig. 10b), and a relatively new type of system used in plasma-source ion implantation (PSII) (Fig. 10c).

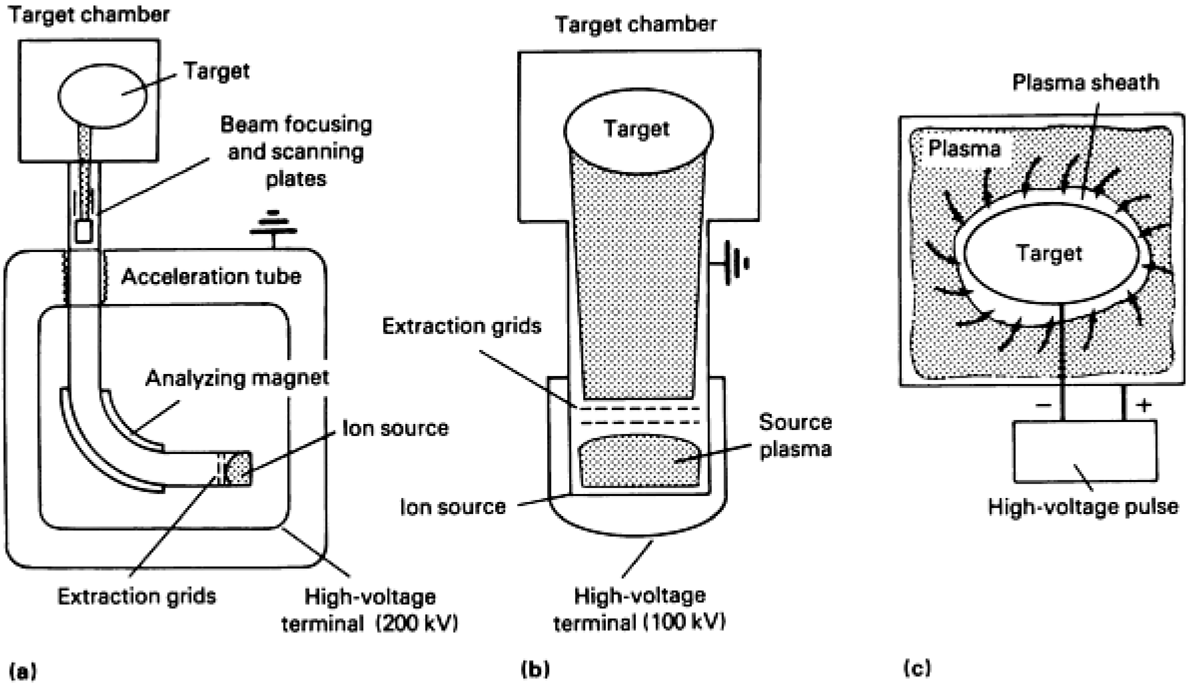


Fig. 10 Schematics of ion implantation equipment. (a) Mass-analyzed implantation. (b) Broad-beam implantation. (c) Plasma-source implantation

Ion Source and Ion Extractor. Ions are generated in the ion source. Ionization of gaseous elements, such as nitrogen, oxygen, or carbon (from methane, CH_4), is typically accomplished by introducing the gases into a source at moderate vacuum and bombarding them with energetic electrons. The electrons are supplied from an external source (such as a heated filament) or from a plasma generated from the working gas (N_2, O_2, or CH_4) by applying direct current, radio frequency, or microwave power to the source. For beams of elements that normally exist in a solid state, a volatile compound (for example, BCl_3 for boron, or a compound formed by passing a halide gas over heated metal) can be used as the working gas. Alternatively, a molten pool of metal can be generated in the source to provide partial pressure of elements (for example, titanium or copper), or physical sputtering processes can be built into the source to generate a plasma of the desired element that can be ionized by electron bombardment. The plasma is typically confined in the source by magnetic fields. Charged ions are extracted from the source by an ion extractor, which typically consists of a series of electrically biased grids with a number of apertures.

Analyzing Magnet. Because some source designs produce unwanted ions (as in the case of sputter sources that use inert gases such as argon as the working gas, thus producing argon ions in addition to the desired ion), they typically include an analyzing magnet (Fig. 10a), which selects the desired ions from the beam of ions extracted from the source. In other designs (Fig. 10b), a high-purity source of vapor exists, making it unnecessary to mass analyze the extracted beam (unless selection of a particular charge state of the ions is desired). Recent advances in broad-beam source technology indicate that high-current ion sources for both gaseous and metallic ion beams can be achieved (Ref 57, 58, 59).

Acceleration Tube and Target Chamber. The ion source, extraction grids, and analyzing magnet are typically floated at voltages ranging from 25 to several hundred kV for most commercial implantation systems. The next component in the systems shown in Fig. 10(a) and 10(b) is an acceleration tube, which accelerates the ions to ground potential.

After exiting the acceleration tube, the ion beam is focused, steered, and rastered to achieve the desired beam shape and current density. The last component of the implantation system is a target chamber, consisting of a vacuum enclosure and equipment for manipulating the parts to be implanted in the beam of energetic ions. Mechanical motion of the substrates coupled with beam rastering generally is used to obtain the desired doses. Beam rastering is not used in broad-beam implanters. In some instances, masks must be placed in front of parts with complex shapes to limit implantation at off-normal directions and thus minimize sputtering effects.

System Differences. The major difference between the systems illustrated in Fig. 10(a) and (b) involves accelerating voltages and beam current. Broad-beam implantation systems typically operate at acceleration voltages up to 100 kV and can deliver large beam currents (in excess of 500 mA for gaseous ions and 50 mA for metallic ions) on target. Mass-analyzed implanters designed for commercial applications typically operate at acceleration voltages up to 200 to 400 kV, with beam currents up to 10 to 20 mA on target. Another difference is the size of the beam on target. As the name implies, the size of the ion beam on target for broad-beam implanters can be quite large (in the range of 0.5 m, or 1.5 ft, diam), while mass-analyzed beams are typically on the order of 10 mm (0.4 in.) diam. As a result, mass-analyzed beams are usually rastered over the component to be implanted, whereas broad-beam ion beams are not.

The PSII system illustrated in Fig. 10(c) is relatively new. Rather than extracting and accelerating an ion beam from a source and impinging the beam on the part in a target chamber, the PSII process places the part inside a plasma and repetitively pulses the part to be implanted at high negative voltages (typically -100 kV) to obtain a flux of energetic plasma ions on it (Ref 60). Because the plasma surrounds the part, and because the ions are accelerated normal to the part surface, the plasma-source implantation occurs over the entire surface, thereby eliminating the need to manipulate the part in front of the ion beam as in Fig. 10(a) and (b).

In addition to the basic components illustrated in Fig. 10, most implanters have equipment for monitoring and controlling the ion beam. Such diagnostic equipment typically consists of components for monitoring the ion dose on target, current density and beam shape, and substrate temperature, and for cooling and heating the substrate.

Applications of Ion Implantation

Commercial use of ion implantation for tribological applications dates back to the mid-1970s. Many applications are related to improvements in tooling used in metalforming and cutting operations. The initial cost of the tooling and increased productivity due to decreased downtime for changing tools easily offset the added cost of implantation. In these applications, use of implanted tools results in longer tool life, improved surface finishes of worked material, reduced pickup or transfer of material, easier release from molds, and improved corrosion resistance. End-use applications of ion implantation continue to grow in areas where the components are small and the implantation cost per part is low, where replacement cost is high, or where long shelf life (corrosion resistance) is required.

Selection of the dose and the type of ion to be implanted depends greatly on the tribological conditions. These conditions include not only the composition of the components but also the environment, operating temperature, and stresses. In most cases, the selection of the implantation conditions is empirically determined. Table 2 lists some of the ions that improve wear (adhesive/abrasive, fatigue, and corrosive) and friction properties of different alloys.

Table 2 Elements implanted to improve wear and friction properties

Alloy	Implantation element				Ref
	Adhesive and abrasive wear	**Fatigue wear**	**Corrosion and oxidation wear**	**Friction**	
Aluminum alloys	N		Mo		50, 61
Beryllium alloys	B				50, 61, 62
Ceramics	N, O, Cr, Zr, Y	N, O			61
Cemented carbides (WC)	N, Co				61, 62
Copper alloys	B, Ti	Al	Al, Cr		50, 61, 62, 63
Steels	N, Ti + C, Y, Ta + C	B, N, Ti	Cr, Ta, Y	Ti, Ti + C, Ag, Sn, Au	50, 52, 61, 62, 63
Superalloys	C, N, Y, Ti + C	C	Al, Y, Ta, Pt, Au, Ce		50, 61, 62
Titanium alloys	B, C, N	C, N, Al	C, N, Pd	Ag, Sn	50, 61, 62, 63
Zirconium alloys	C, N, Cr + C		Cr, Sn		61, 62

Most implantations for tribological applications where chemical rather than microstructural changes are required use doses in the range of 10^{17} to 10^{18}ions/cm^2. Under typical implantation conditions at nominal room temperature, sputtering limits the maximum concentration of implanted species that can be achieved. Thus, implanting to doses much greater than approximately 1×10^{18}ions/cm^2 (depending on the sputtering yield) is largely useless.

An exception to this dose limit can occur if the implantation is performed at elevated temperatures or at ultrahigh current densities that produce high substrate temperatures or that induce radiation-enhanced diffusion and thus permit mobile implanted species to diffuse deeper into the substrate. As an example, broad-beam implantation of type 304 stainless steel (Ref 31) with 60 keV nitrogen ions to a dose of 1×10^{19} ions/cm^2 at an ultrahigh current density of 500 μA/cm^2 (3.2 mA/in.^{2}) in a pulsed manner to control the substrate temperature to approximately 500 °C (930 °F) resulted in a nitrogen concentration level of 15 to 20 at.% that extended to depths greater than 10 μm. This is approximately 100 times deeper than the projected range (Fig. 3a) of 60 keV nitrogen ions in iron, indicating that thermal and/or radiation-enhanced diffusion allowed the nitrogen to penetrate farther into the substrate. In contrast, implantation at lower current densities (100 μA/cm^2, or 645 μin.^{2}) resulted in lower substrate temperatures (approximately 380 °C, or 715 °F), and in concentration profiles that more closely matched those predicted by athermal calculations.

For typical ion implantation procedures performed at nominal room temperature, Table 3 lists approximate ion-dose regimes where different hardening mechanisms are expected to be first observed. For conventional cold-working mechanisms, it is evident that heavy ions are more effective, because they displace atoms much more readily than light ions. For the other mechanisms, a certain concentration of foreign species is required, and thus higher implantation doses must be used to achieve the desired result.

Table 3 Approximate dose regimes for effective hardening during ion implantation

Hardening mechanism	Dose, ions/cm^2	
	Light ions (C, N)	**Heavy ions**
Cold working (defect/dislocation tangles)	10^{14}-10^{16}	10^{13}-10^{16}
Interstitial solution hardening	10^{16}-10^{17}	. . .
Substitutional solution hardening	. . .	10^{16}-10^{17}
Second-phase hardening	10^{17}-10^{18}	10^{16}-10^{17}
Phase transformation	10^{17}-10^{18}	10^{17}-10^{18}

Specific examples of applications of ion implantation in metalforming and cutting are given in Table 4, and Table 5 lists examples of end-use applications. In many of the applications noted in Tables 4 and 5, the implanted components generally operate at relatively low temperatures, where microstructural and chemical changes caused by implantation are stable. Thus, operating conditions, particularly the nominal bulk temperature and flash temperature of the asperities, can affect the viability of a particular implantation treatment for improving the tribological performance of a given component. A treatment that works in one situation to reduce adhesive wear may not work in another situation because of different tribological conditions. As a result, it is crucial to understand the tribological environment (loads, stresses, bulk

and flash temperatures, and material properties) and how that environment can influence the microstructural and chemical changes induced by implantation processes.

Table 4 Examples of ion implantation in metalforming and cutting applications

Part	Part material	Process	Work material	Wear mechanism	Ion	Energy, keV	Benefit	Ref
Tool inserts	TiN-coated tool steel	Machining	4140	Abrasion	N	80	3× life	63
	TiN-coated WC	Machining	321 SS	Abrasion	N	80	2× life	63
	TiN-coated WC	Machining	4140	Abrasion	N	50	5× life	64
Taps	HSS	Tapping	4140	Abrasion	N	80	3× life	63
	HSS	Tapping	4130	Abrasion	N	80	5× life	63
	HSS	Tapping	4140	Abrasion	N	50	10× life	64
	M35	Tapping	. . .	. . .	N_2	200	4× life	65
	M7	Tapping	. . .	. . .	N	100	2× life	65
Cutting blade	M2	Cutting	1050	. . .	N	100	2× life	66
	M2	Cutting	SAE 950	. . .	N	100	4× life	66
Dies	D2	Forming	321 SS	Abrasion	N	80	2× life	63
	M2	Forming	Steel	. . .	N	100	2-12× life	66
	M2	Forming	1020	. . .	N	100	Negligible effect	66
	D6	Forming	TiO_2 and rubber	. . .	N	100	6× life	65
	WC	Forming	Aluminum	. . .	N	100	2× life	65
	Steel	Forming	Tin cans	. . .	N	100	2-6× life	65
Molds	Aluminum and brass	Forming	Polymers	. . .	Ni, P	. . .	Better release	67
	Chromium/D2	Forming	Polymers	Adhesion	N	50	5× life	64
Rollers	H13	Rolling	Steel	. . .	N	100	5× life	65

Table 5 Examples of end-use applications of ion implantation

Part	Part material	Application/ environment	Mating material	Wear mechanism	Ion	Energy, keV	Benefit	Ref
Bearing rings	440C SS	Shuttle engine	440C SS	Rolling-contact fatigue and corrosive wear	Cr + N	. . .	Improved corrosive wear resistance	67
Ball bearings	440C SS	Shuttle engine	440C SS	Rolling-contact fatigue and corrosive wear	Ti + C	. . .	Lower friction (30×) and wear	64
			440C SS		Ti	180		67
Roller bearings	52100	Oil	9310 steel	Corrosive wear	Cr	180	Improved corrosive wear resistance	64
Ball valves	Titanium alloy	Corrosive liquid	Polymer	Rubbing	N	. . .	Improved corrosive wear resistance	67
Reciprocating pistons	Tool steel	Cryogenic liquid	Steel	Sliding	Ti + C	. . .	. . .	67
Co-Cr homogenizer	Co-Cr alloy	Milk	Co-Cr alloy	Cavitation	Ti	180	Improved life	64
Dental tools	Titanium alloy	Saliva	Bone/teeth	Fretting	N	. . .	. . .	67
					N	100	2× life	65
Prosthetics	Titanium alloys	In vivo	Polymers/bone	Corrosive wear	N	. . .	Increased life	67
					N	50	Delay of corrosion	64

References

1. G.H. Kinchin and R.S. Pease, The Displacement of Atoms in Solids by Radiation, *Rep. Prog. Phys.,* Vol 18,

1955, p 1-51

2. J.F. Ziegler, J.P. Biersack, and U. Littmark, *The Stopping and Range of Ions in Solids,* Vol 1, Pergamon Press, 1975
3. H.H. Anderson, *The Stopping and Range of Ions in Solids,* Vol 2, *Bibliography of Experimental Range and Stopping Power Data,* Pergamon Press, 1977
4. H.H. Anderson and J.F. Ziegler, *The Stopping and Range of Ions in Solids,* Vol 3, *Hydrogen Stopping Powers and Ranges in All Elements,* Pergamon Press, 1977
5. J.F. Ziegler, *The Stopping and Range of Ions in Solids,* Vol 4, *Helium Stopping Powers and Ranges in All Elements,* Pergamon Press, 1977
6. J.F. Ziegler, *The Stopping and Range of Ions in Solids,* Vol 5, *Handbook of Stopping Cross-Sections for Energetic Ions in All Elements,* Pergamon Press, 1980
7. U. Littmark and J.F. Ziegler, *The Stopping and Range of Ions in Solids,* Vol 6, *Handbook of Range Distributions for Energetic Ions in All Elements,* Pergamon Press, 1980
8. D.K. Brice, *Ion Implantation Range and Energy Deposition Distributions,* Vol 1, *High Incident Ion Energies,* Plenum, 1975
9. K.B. Winterbon, *Ion Implantation Range and Energy Deposition Distributions,* Vol 2, *Low Incident Ion Energies,* Plenum, 1975
10. J. Lindhard, V. Nielsen, M. Scharff, and P.V. Thomsen, Integral Equations Governing Radiation Effects (Notes on Atomic Collisions, III), *Mat. Fys. Medd. Dan. Vid. Selsk,* Vol 33 (No. 10), 1963, p 1-42
11. J.F. Ziegler, J.P. Biersack, and U. Littmark, *The Stopping and Range of Ions in Solids,* Vol 1, Pergamon Press, 1975, p 109-141
12. M.T. Robinson and I.M. Torrens, Computer Simulation of Atomic-Displacement Cascades in Solids in the Binary-Collision Approximation, *Phys. Rev. B,* Vol 9 (No. 12), 1974, p 5008-5024
13. S.N. Bunker and A.J. Armini, Modeling of Concentration Profiles From Very High Dose Ion Implantation, *Nucl. Instrum. Methods Phys. Res.,* Vol B39, 1989, p 7-10
14. P. Sigmund, Theory of Sputtering. I. Sputtering Yield of Amorphous and Polycrystalline Targets, *Phys, Rev.,* Vol 184 (No. 2), 1969, p 383-416
15. N. Matsunami *et al.,* Energy Dependence of the Ion-Induced Sputtering Yields of Monatomic Solids, *Atomic Data and Nuclear Data Tables,* Vol 31, 1984, p 1-80
16. G.R. Fenske, "Helium-Induced Blistering and Volume Swelling in Nickel," Ph.D. thesis, University of Illinois at Urbana-Champaign, 1979
17. P.R. Okamoto and M. Meshi, Solid-State Amorphization-Particle Irradiation, *Science of Advanced Materials,* H. Weidersich and M. Meshi, Ed., ASM International, 1990
18. D.F. Pedraza and P.J. Maziasz, The Effects of Microstructure in the Morphology and Kinetics of Amorphization Induced by Ion Irradiation, *Materials Research Society Symposium Proceedings,* Vol 157, 1990, p 783-788
19. E. Johnson, Martensitic Transformations in Ion Implanted Stainless Steels, *Materials Research Society Symposium Proceedings,* Vol 157, 1990, p 759-770
20. C.J. McHargue, P.S. Sklad, P. Angelini, C.W. White, and J.C. McCallum, Structure of Amorphous Al_2O_3 Produced by Ion Implantation, *Materials Research Society Symposium Proceedings,* Vol 157, 1990, p 505-512
21. L.E. Rehn, S.T. Picraux, and H. Wiedersich, Overview of Surface Alloying by Ion, Electron, and Laser Beams, *Surface Alloying by Ion, Electron, and Laser Beams,* L.E. Rehn, S.T. Picraux, and H. Weidersich, Ed., American Society for Metals, 1987
22. R.S. Averback and M.A. Kirk, Atomic Displacement Processes in Ion Irradiated Materials, *Surface Alloying by Ion, Electron, and Laser Beams,* L.E. Rehn, S.T. Picraux, and H. Wiedersich, Ed., American Society for Metals, 1987
23. D.M. Follstaedt and S.T. Picraux, Microstructures of Surface-Melted Alloys, *Surface Alloying by Ion, Electron, and Laser Beams,* L.E. Rehn, S.T. Picraux, and H. Weidersich, Ed., American Society for Metals,

1987

24. C.R. Clayton, Chemical Effects of Ion Implantation: Oxidation, Corrosion, and Catalysis, *Surface Alloying by Ion, Electron, and Laser Beams,* L.E. Rehn, S.T. Picraux, and H. Weidersich, Ed., American Society for Metals, 1987
25. B. Rauschenbach and V. Heera, Amorphisation of Metals by Ion Implantation and Ion Beam Mixing, *Defect and Diffusion Forum,* Vol 57-58, 1988, p 143-188
26. J.S. Custer, M.O. Thompson, D.C. Jacobson, and J.M. Poate, Interface Velocity and Noble Metal Segregation During Ion Beam Induced Epitaxial Crystallization, *Materials Research Society Symposium Proceedings,* Vol 157, 1990, p 113-118
27. T.K. Chaki, Mechanism of Ion-Induced Solid-Phase Crystallization and Amorphization, *Materials Research Society Symposium Proceedings,* Vol 157, 1990, p 173-178
28. A. Arnold, R. Ramlau, and E.H. Kaat, Cross-Sectional TEM Investigations of Iron After High-Dose Nitrogen Implantation, *Nucl. Instrum. Methods Phys. Res.,* Vol B59/60, 1991, p 726-730
29. G. Marest, Nitrogen Implantation in Iron and Steels, *Defect and Diffusion Forum,* Vol 57-58, 1988, p 273-326
30. K. Hohmuth, B. Rauschenbach, A. Kolitsch, and E. Richter, Formation of Compounds by Metalloid Ion Implantation in Iron, *Nucl. Instrum. Methods Phys. Res.,* Vol 209/210, 1983, p 249-257
31. R. Wei, P.J. Wilbur, O. Ozturk, and D.L. Williamson, Tribological Studies of Ultra-high Dose Nitrogen-Implanted Iron and Stainless Steel, *Nucl. Instrum. Methods Phys. Res.,* Vol B59/60, 1991, p 731-736
32. D.L. Williamson, R. Wei, and P.J. Wilbur, Effects of Rapid, High-Dose, Elevated Temperature Ion Implantation on the Microstructure and Tribology of Ferrous Surfaces, *Nucl. Instrum. Methods Phys. Res.,* Vol B56/57, 1991, p 625-629
33. E. Johnson, Ion Implantation Induced Martensitic Transformation in Metals, *Defect and Diffusion Forum,* Vol 57-58, 1988, p 241-271
34. S. Fayeulle, Ion Implantation in Stainless Steels: Microstructure and Mechanical Properties, *Defect and Diffusion Forum,* Vol 57-58, 1988, p 327-358
35. K. Kobs *et al.,* Strong Wear Reduction of High-Dose Carbon-Implanted AISI 52100 Steel, *Nucl. Instrum. Methods Phys. Res.,* Vol B59/60, 1991, p 746-750
36. J. Bentley, L.D. Stephenson, R.B. Benson, Jr., P.A. Parrish, and J.K. Hirvonen, Second Phase Formation in Aluminum Annealed After Ion Implantation With Molybdenum, *Materials Research Society Symposium Proceedings,* Vol 27, 1984, p 151-156
37. K.V. Jata, D. Janoff, and E.A. Starke, Jr., Modulated Structures in Ion Implanted Al-Fe System, *Materials Research Society Symposium Proceedings,* Vol 27, 1984, p 157-162
38. P.S. Sklad, P. Angelini, M.B. Lewis, and C.J. McHargue, Microstructural Development of TiB_2 Ion Implanted With 1 MeV Nickel, *Materials Research Society Symposium Proceedings,* Vol 27, 1984, p 407-412
39. R.G. Vardiman, Wear Improvement in Ti-6Al-4V by Ion Implantation, *Materials Research Society Symposium Proceedings,* Vol 27, 1984, p 699-704
40. D.R. Olander, "Fundamental Aspects of Nuclear Reactor Fuel Elements," TID-26711-P1, National Technical Information Service, 1976
41. J. Gittus, *Irradiation Effects in Crystalline Solids,* Applied Science, London, 1978
42. L.K. Mansur, Mechanisms and Kinetics of Radiation Effects in Metals and Alloys, *Kinetics of Nonhomogeneous Processes,* G.R. Freeman, Ed., John Wiley & Sons, 1987, p 377-464
43. S.M. Myers and D.M. Follstaedt, Solute Trapping in Ion-Implanted Metals, *Surface Alloying by Ion, Electron, and Laser Beams,* L.E. Rehn, S.T. Picraux, and H. Wiedersich, Ed., American Society for Metals, 1987
44. N.Q. Lam and G.K. Leaf, Mechanisms and Kinetics of Ion Implantation, *J. Mater. Res.,* Vol 1 (No. 2), 1986, p 251-267
45. A.M. Yacout, N.Q. Lam, and J.F. Stubbins, Near-Surface Compositional Modifications of Ternary Alloys

During Sputtering at Elevated Temperatures, *Nucl. Instrum. Methods Phys. Res.,* Vol B42, 1989, p 49-60

46. N.Q. Lam and H. Wiedersich, Bombardment-Induced Segregation and Redistribution, *Nucl. Instrum. Methods Phys. Res.,* Vol B18, 1987, p 471-485
47. N.Q. Lam, Mechanisms and Kinetics of Ion Beam-Induced Compositional Modifications, *Materials Research Society Symposium Proceedings* Vol 100, 1988, p 29-43
48. N.Q. Lam, Ion Bombardment Effects on the Near-Surface Composition During Sputter Profiling, *Surface Interface Anal.,* Vol 12, 1988, p 65-77
49. D.G. Teer and R.D. Arnell, Wear, *Principles of Tribology,* J. Halling, Ed., Macmillan, 1975
50. G.K. Hubler, Surface Alloying by Ion Beams: Mechanical Effects, *Surface Alloying by Ion, Electron, and Laser Beams,* L.E. Rehn, S.T. Picraux, and H. Wiedersich, Ed., American Society for Metals, 1987
51. E. Rabinowicz, Determination of Compatibility of Metals Through Static Friction Test, *ASLE Trans.,* Vol 14, 1971, p 198-205
52. G. Dearnaley, Adhesive, Abrasive and Oxidative Wear in Ion-Implanted Metals, *Mater. Sic. Eng.,* Vol 69, 1985, p 139-147
53. G. Dearnaley, Ion Beam Modification of Metals, *Nucl. Instrum. Methods Phys. Res.,* Vol B50, 1990, p 358-367
54. K.G. Stephens, An Introduction to Ion Sources, *Ion Implantation Science and Technology,* 2nd ed., J.F. Ziegler, Ed., Academic Press, 1988
55. I.G. Brown, *The Physics and Technology of Ion Sources,* Wiley-Interscience, 1989
56. A.T. Forrester, *Large Ion Beams,* Wiley-Interscience, 1989
57. P.J. Wilbur and L.O. Daniels, The Development and Application of an Ion Implanter Based on Ion Thruster Technology, *Vacuum,* Vol 36 (No. 1-3), 1986, p 5-9
58. P.J. Wilbur and R. Wei, "High-Current-Density Metal-Ion Implantation," presented at 4th International Conference on Ion Sources (Bensheim, Germany), 30 Sept-4 Oct 1992, to be published in *Rev. Sci. Instrum.*
59. I.G. Brown, J.E. Galvin, and R.A. MacGill, Broad-Beam Multi-Ampere Ion Source, *Rev. Sci. Instrum.,* Vol 61 (No. 1), 1990, p 577-579
60. J.R. Conrad, Plasma Source Ion Implantation: A New Approach to Ion Beam Modification of Materials, *Mater. Sci. Eng. A,* Vol 116, 1989, p 197-203
61. P. Sioshansi, Surface Modification by Ion Implantation, *Mach. Des.,* Vol 58 (No. 6), 1986, p 61-65
62. P. Sioshansi, Tailoring Surface Properties by Ion Implantation, *Mater. Eng.,* Vol 104 (No. 2), 1987, p 19-23
63. A. Golzales, U.S. Army Aviation Systems Command Depot Engineering & RCM Support Office, Corpus Christi Army Depot, private communication
64. A.J. Armini, Implant Sciences Corp., private communication
65. C.A. Straede, Practical Applications of Ion Implantation for Tribological Modification of Surfaces, *Wear,* Vol 130, 1989, p 113-122
66. R.B. Alexander, Industrial Wear Testing of Metal Stamping Tools Implanted With Nitrogen: Trends and Results, *Ion Implantation and Plasma Assisted Processes,* ASM International, 1988, p 17-20
67. R.W. Oliver, Spire Corp., private communication

Laser Surface Processing

Khershed P. Cooper, Naval Research Laboratory

Introduction

HIGH-POWER LASERS have been used to process materials to improve their wear resistance since the late 1960s. Laser surface processing encompasses established techniques, such as laser transformation hardening, which has achieved production-level capabilities, and techniques that are still in a development stage, such as laser cladding, which is emerging as a viable alternative to other hard-facing processes, such as plasma deposition.

The laser beam is a chemically clean light source that delivers a precisely controlled quantity of energy to localized regions. By this means, only those areas exposed to wear are processed for improvement in wear resistance. The laser beam is easily maneuvered by optical elements and can be adapted to automation. High power densities and low interaction times result in rapid heating and cooling, which produces a shallow heat-affected zone, low distortion of the workpiece, and minimal deterioration of bulk properties. Rapid solidification effects produce refined and novel microstructures that result in improved properties.

Laser Characteristics. Most laser surface processing utilizes continuous wave (CW) lasers. In limited instances, a pulsed laser is used, for example, to harden the tip of a tool bit. The most common lasing medium for materials processing is CO_2 gas. The 10.6 μm wavelength CO_2 laser couples poorly with metals, but generally has a high average power output between 1 and 25 kW. With the help of radiation-absorbing agents, metals with high melting points can be processed. The excitation-mode structure of the CO_2 laser is usually "doughnut-shaped," which, after being modified by optics, produces a near-Gaussian energy distribution on the sample surface.

In its simplest form, laser-material interaction involves the heating and melting of a moving substrate with a stationary beam. This "spot" energy source is typically 1 to 2 mm (0.04 to 0.08 in.) in diameter and produces a melt pass with a width from 2 to 3 mm (0.08 to 0.12 in.). By overlapping several of these passes, usually by 25 to 50%, larger areas can be processed. In addition to being slow, this technique can often cause tempering and cracking within the overlap regions, thereby lowering the quality of the product surface.

An alternative method is the use of beam-shaping optics to produce broad energy distributions on the sample surface. Scanning the beam over the substrate, either by oscillating a flat mirror or by rotating a polygonal mirror, produces a "line" energy source. An integrated beam, which is formed by an integrated mirror, produces a "rectangular" energy source. When interacting with a moving workpiece, either energy source produces a strip of modified surface. The use of beam-shaping optics represents an efficient method of achieving the hardening of selected areas on the workpiece, which is required by most wear applications.

In addition to shaping the beam, optical elements can be designed and arranged to direct it onto the workpiece in many ways. For example, a reflecting mirror can be used to project the laser beam onto the inner wall of a rotating tube. Toric mirrors also can be designed to form a continuous band of radiation to treat inner or outer surfaces of cylinders and cones. Prisms can also be used to split the laser beam and direct the two parts at different angles to a shaped surface, such as a gear groove.

Laser surface modification techniques that are used to improve the wear resistance of metals include transformation hardening, melting, alloying, cladding, and particle injection. The processing conditions, examples of microstructures, wear characteristics, and applications associated with each technique are discussed in detail in this article.

Parameters for laser surface processing depend on the type of modification (for example, transformation hardening or cladding) and thermophysical properties of the material (such as thermal diffusivity or heat capacity). The important laser processing parameters are power density and interaction time, which are functions of the average laser power, the sample translation rate, and the beam diameter. Rapid melting or glazing requires either high power densities and short interaction times or high power levels and high translation rates, whereas transformation hardening requires either low power densities and long interaction times or low power levels and low translation rates.

Other processing variables, such as the composition of the atmosphere or the rate of material addition, depend on the particulars of the process. For example, consideration must be given to the need to protect against oxidation and the desire to have certain characteristics of the surface layer, for example, thickness, composition, and microstructure. Depending on the processing conditions, surface layer thicknesses can range from 0.01 to 5 mm (0.0004 to 0.2 in.). The longer the interaction time, the deeper the treated layer will be.

Processing parameters for cladding and surface melting typically modify the surface to depths from 0.5 to 5 mm (0.02 to 0.2 in.) with microstructures that are similar to welded structures. The parameters for rapid surface melting or glazing modify surfaces from 0.02 to 0.6 mm (0.0008 to 0.02 in.) deep, but the resulting rapid quench rates from 10^5 to 10^7 K/s allow for the synthesis of novel microstructures. Material addition practices will depend on the specific laser surface processing technique under consideration. Steen (Ref 1) has summarized the process variables and their methods of measuring (Table 1).

Table 1 Process variables and methods of measuring

Process variable	Method of measuring
Laser power	Flowing cone calorimeter
Beam diameter	Rotating rod
Beam mode structure	Acrylic prints, fluorescent screens
Sample speed	Electronic timer
Material addition flow rate	Load cell on feed hopper, acoustic emission sensor on feed line
Particle velocity	Rotameters, fast-action photography, photographic tracer technique

Laser Transformation Hardening

Processing. The wear resistance of highly stressed ferrous alloy components, such as gears and bearings, can be improved by transformation hardening. Because ferrous materials are very good heat conductors, the high heat fluxes generated by lasers are most suitable to heat the surface layer to austenization levels without affecting the bulk temperature of the sample. The ensuring self-quenching is rapid enough to eliminate the need for external quenching to produce the hard martensite in the heated surface. Because ferrous alloys have high reflectivity, absorptive coatings such as manganese phosphate and graphite are applied to the workpiece for efficient laser heating. Most of these coatings burn off and normally do not affect the microstructure. Processing conditions for laser transformation hardening are typically power densities that range from 5 to 100 MW/m^2 (3.2 to 64.5 kW/in.^{2}) and interaction times ranging from 0.01 to 10 s. Usually, inert gas shielding is used.

A rectangular beam is often used for laser transformation hardening. According to Sandven (Ref 2), the case depth depends on the hardenability of the material and rarely exceeds 2.5 mm (0.1 in.). For example, in medium-carbon steel, a case depth of 1.25 mm (0.05 in.) is common, whereas in mild steel, a case depth of only 0.25 mm (0.01 in.) is possible for the same processing conditions. Because of the rapid heating and cooling rates, steels that are not hardenable with other heat treatment methods can be hardened with laser processing.

The hardened case depth and the transition zone also depend on the coarseness of the initial microstructure. With a coarse starting structure, the case depth is shallower and the transition from full martensite hardness to base metal hardness is more gradual than with a fine starting structure. According to Seaman (Ref 3), coarser structures provide longer diffusion paths in which to achieve uniform austenization. To facilitate complete transformation during laser transformation hardening, materials with fine starting microstructures, such as closely spaced pearlite, are recommended. In many cases, the part is conventionally quenched and tempered prior to laser processing.

Because laser processing can occur too rapidly to allow for complete diffusion of carbon in the austenite, the interaction times are kept longer. The laser power density is kept low to prevent surface melting, which can form harmful eutectic carbides and can blunt the sharp edges and corners of a complex workpiece. Melting can also change the surface topography, requiring postmachining. The careful design of the beam shape and energy distribution can minimize surface melting problems. During closure of a circular hardening pass or during overlapping of several passes, back-tempering can occur in the overlapped regions. If this is detrimental, then the overlapped area should be located in a position of low service wear.

For example, the interior walls of internal combustion engines can be hardened by spiral laser passes, which results in some unhardened regions without causing harmful effects. It has been claimed that the use of a laser-hardened spiral on a shaft that runs in a bearing results in one-third of the wear that is experienced when using a uniformly induction-hardened shaft, because, in the former case, the wear debris becomes safely embedded in the softer interspiral regions (Ref 4). In one line-hardening procedure for plain carbon steels, the laser-heated areas rise above the sample surface and act as hard bearing tracks in a tough material, because of the volume expansion of the martensite (Ref 5).

In laser transformation hardening, the thermal energy is generated by absorption of the laser radiation at the surface. The increase in temperature below the surface is by conduction only. If the rate of heat absorption and the thermal properties of the sample are known, then the temperature distribution in the sample can be calculated. By this means it is possible to determine the optimum laser processing conditions.

Heat flow expressions are available in the literature to make these calculations (Ref 2). The simplest expression assumes heat flow to be one-dimensional, normal to the surface. In reality, heat diffuses in all directions and is lost by radiation from the surface. Furthermore, the thermal properties are temperature dependent, and the workpiece is not always large enough to be considered an infinite heat sink. These factors need to be considered when determining the temperature distributions. However, for high processing speeds, thick samples, and large beam sizes, the one-dimensional model gives good estimates. Using the one-dimensional heat flow model, cooling rates can also be calculated. When used in conjunction with cooling transformation diagrams, the adequacy of the self-quenching for hardening can be predicted.

Examples of Laser-Hardened Surface Microstructures. In ferrous alloys, the formation of martensite is essential for hardening and increased wear resistance. In laser-hardened carbon steels, such as 1018 and 1045, the relative amounts of ferrite and martensite depend on the carbon content. In laser-hardened alloy steels, such as 4340, 8620 and 52100, a very fine martensite forms (Ref 6). Courtney and Steen (Ref 7) found that in a 0.36% C steel, the microstructure was dependent on sample translation rate. At low sample speeds, areas of high-carbon martensite formed, whereas at high sample speeds, an inhomogeneous martensite mixed with austenite formed.

In cast irons, laser heat-treated surfaces appear in light contrast, as shown in Fig. 1(a) for gray iron and in Fig. 1(b) for ductile iron (Ref 8). The microstructure of laser-hardened gray cast iron is generally fine martensite that contains flake graphite, as shown in Fig. 2(a) (Ref 8). Acicular bainite can sometimes occur in the martensite matrix (Ref 9). The microstructure of laser-hardened ductile iron is also fine martensite, but contains "bull's eye" graphite nodules, as shown in Fig. 2(b) (Ref 8). A closer examination of Fig. 2(b) reveals a narrow region of martensite between the graphite nodule and the ferrite ring, which indicates the extent of carbon diffusion occurring during processing.

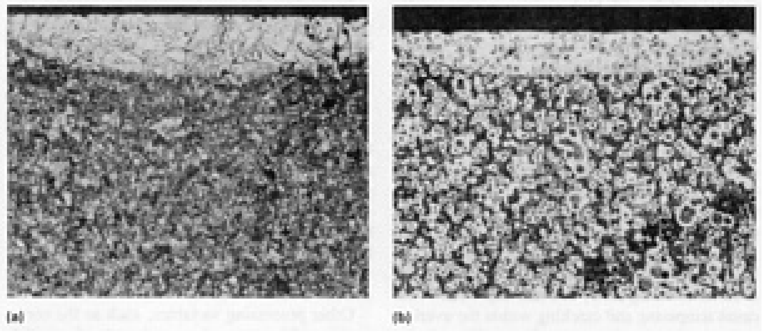


Fig. 1 Cross sections of laser heat-treated surfaces in cast irons. (a) Gray iron. (b) Ductile iron. Source: Ref 8

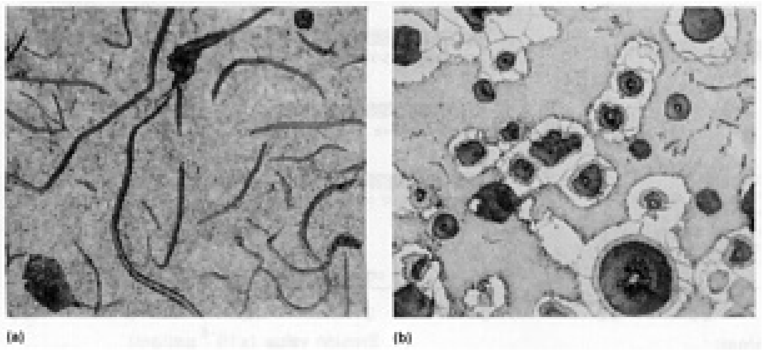


Fig. 2 Microstructures of laser-hardened cast irons. (a) Gray iron. (b) Ductile iron. Source: Ref 8

According to Molian and Baldwin (Ref 10), the extensive case that developed during laser hardening of gray and ductile irons was made up of sequential regions with varying degrees of microstructural modification. The region just below a very thin melt zone contained plate martensite, retained austenite, and graphite, whereas the region just above the base metal contained a mixture of refined martensite and untransformed pearlite.

Wear Behavior of Laser-Hardened Surfaces. Gnanamuthu (Ref 11) reported pin-on-disk wear test results on laser-hardened AISI 1045 steel specimens. After testing for 10 h, the weight loss that was due to wear in the laser-treated samples ranged from 0.6 to 1.4 mg (2.1 to 4.9 × 10^{-5} oz), whereas the weight loss in the untreated samples was 4.8 mg (1.7 × 10^{-4} oz). Using a two-disk machine, Lu and Zhang (Ref 12) found that laser hardening significantly improved the wear resistance in normalized or austempered ductile iron by the formation of strain-induced martensite. From their study of pin-on-disk wear behavior of laser-hardened gray and ductile cast irons, Molian and Baldwin (Ref 10) found an improvement in scuffing and sliding wear resistance with an increase in case depth as shown in Fig. 3. On the same samples, erosive wear test results showed that the erosion rate depended on the surface hardness and the case depth, as shown in Fig. 4, and increased as the matrix microstructure varied from ledeburite to tempered martensite to pearlite (Ref 13).

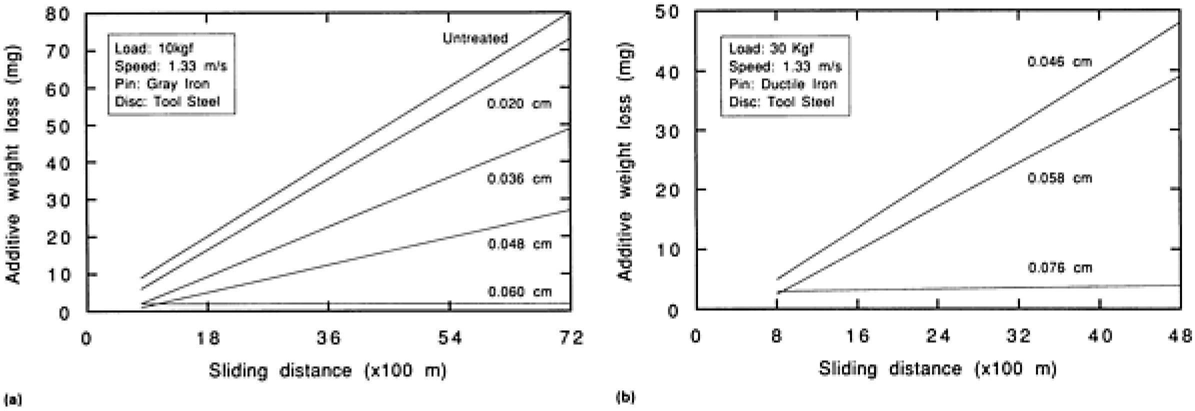


Fig. 3 Sliding wear behavior of laser-hardened cast irons as a function of case depth. (a) Gray iron. (b) Ductile iron. Source: Ref 8

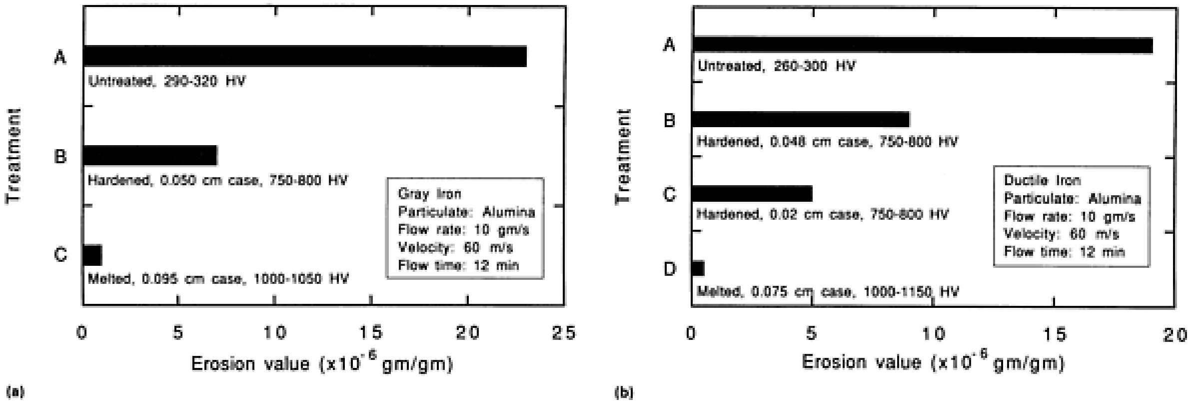


Fig. 4 Erosive wear behavior of laser-hardened cast irons as a function of surface hardness and case depth. (a) Gray iron. (b) Ductile iron. Source: Ref 13

Tomlinson *et al.* (Ref 14) observed that besides reducing wear, laser hardening of flake graphite iron resulted in very little plastic deformation, no graphite deposition on the opposing wheel, and no adhesive damage. Trafford *et al.* (Ref 15) used a reciprocating pad-on-plate technique to evaluate the wear resistance of laser-processed gray iron and found that fully martensitic and surface-melted ledeburite microstructures had the lowest wear rates (Fig. 5).

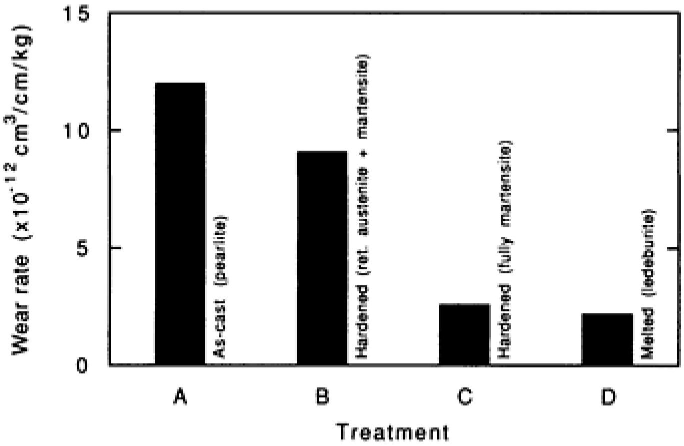


Fig. 5 Abrasive wear rates in gray iron as a function of microstructure arising from various laser-hardening treatments. Source: Ref 15

Applications of Laser Transformation Hardening. Wineman and Miller (Ref 16) describe laser hardening of selected areas on an automotive power steering gear housing made of ferritic malleable cast iron. The bands of highest wear on the housing were laser hardened to 60 HRC by transforming a ferritic-pearlitic-temper carbon microstructure to temper carbon surrounded by martensite. No distortion was measured and wear was reduced by 90%. Previously, the part was nitrided, requiring a costly annealing operation to maintain concentricity. Yessik (Ref 17) describes laser hardening of a 4150 steel gear spline using a long depth of field and a soft focus to ensure simultaneous hardening of tip and root.

Sandven (Ref 2) describes a cast iron cam-shaft lobe that was laser hardened by rotating under a rectangular beam. The cam-shaped lobe was hardened to a constant case depth of 1 mm (0.04 in.) by varying the angular speed from 12.5 mm/s (0.5 in./s) at the cylindrical portion to 3 mm/s (0.12 in./s) at the flat portion. The laser-hardened surface required no postmachining, but induction hardening had been used previously, causing considerable distortion that required postmachining.

Other laser hardening examples given by Sandven (Ref 2) are an SAE 1045 steel gear and a 25 mm (1 in.) diam 4140 steel cylinder. The gear was hardened to 60 HRC by splitting the laser beam using a prism and directing each half to the gear groove in the form of a 12.5 mm (0.5 in.) square shape. Using a toric mirror, a 5 mm (0.2 in.) wide band of laser energy was formed to harden the conical top of the cylinder. The entire cone was treated by translating the workpiece axially. Because the focusing distance varied along the conical top, the laser power was increased from 3.5 kW at the tip to 7.6 kW at the cylindrical portion.

By properly shaping the laser beam, Seaman and Gnanamuthu (Ref 18) describe hardening of only the sides of the grooves in a gray iron piston ring. Melting of the sharp corner was avoided by precisely directing the beam. This is an inaccessible area that would have been difficult to harden by conventional means. The authors also describe laser hardening of the faces of a malleable cast iron spacer to a hardness of 58 HRC. In this case, the laser beam power was ramped down by an aperture to minimize tempering where the circle closed out.

La Rocca (Ref 19) reports laser hardening of gear and rack teeth at an Italian automotive manufacturer. An oscillating beam and beam splitting mirrors were used to harden the two faces of each tooth. Hardened cases of 50 to 55 HRC were obtained on C-45 steel steering racks. At the electromotive division of an American automotive manufacturer, cylinder liners in high-performance turbocharged diesel engines were made more scuff-resistant by using a 10 mm (0.4 in.) diam laser beam (Ref 20). The entire surface was treated by continuously rotating the component and overlapping several paths. No heat treating had been performed previously, and laser hardening improved the service life.

Laser Melting

Processing. Laser melting requires higher power densities than the levels used for laser transformation hardening. The workpiece is often made absorptive either by using coatings similar to those used for laser heating or by increasing surface roughness, for example, by sand blasting. Laser melting can harden alloys that cannot be hardened by laser transformation hardening. In ferritic malleable gray iron, melting enhances the diffusion of carbon, and the ensuing rapid quench produces a hardened region.

One problem with laser melting is that once the melt forms, the absorptive coating is removed, and the melt pool becomes reflective. Water-cooled, optical feedback systems described by Steen (Ref 21) can be used to increase the coupling efficiency. An oscillating beam can be employed to melt larger areas, which makes the mode structure of the beam less important.

According to Bergmann (Ref 22), the production of a crack-free, oxidation-free, and porosity-free layer requires special arrangements for the shielding gas, which allow for changes in the gas flow rate, angle of impingement, and shape of the gas jet. Proper direction of the gas jet can prevent massive plasma formation, which is often responsible for the formation of cracks, as well as oxidation. Processing conditions for laser melting are typically a power density from 10 to 3000 MW/m^2 (6.5 to 1935 $kW/in.^2$) and an interaction time from 0.01 to 1 s. Inert gas shielding is used to prevent oxidation of the surface.

Examples of Laser-Melted Surface Microstructures. Microstructural changes with laser melting are in the forms of grain refinement, solid solutions, and fine dispersions of precipitates. All of these can contribute to the hardening and strengthening of the surface. Laser-melted surfaces of cast irons appear dendritic, as shown in Fig. 6(a) for gray iron and in Fig. 6(b) for ductile iron (Ref 8). Below the melt zone is the heat-affected zone, which appears in lighter contrast in Fig. 6.

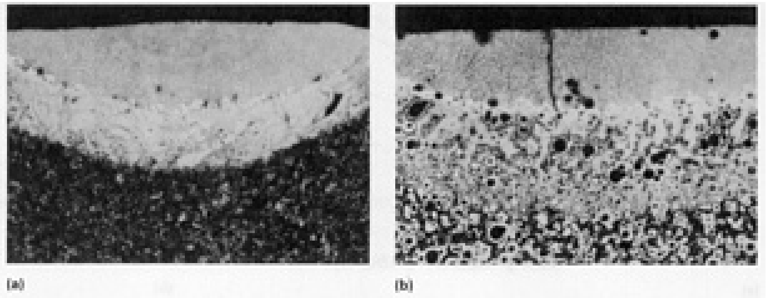


Fig. 6 Cross sections of laser-melted cast iron surfaces. (a) Gray iron. (b) Ductile iron. Source: Ref 8

In the solidified melt in cast irons, a ledeburite (mixture of austenite and cementite) structure generally forms (Ref 21, 22). Hardening is caused by graphite dissolution to form cementite and austenite transformation to martensite. Molian and Baldwin (Ref 10) described the formation of predominantly dendritic ledeburite with small amounts of plate-like, high-carbon martensite and retained austenite in the melt zone in gray and ductile irons.

Chen *et al.* (Ref 23) determined that microstructures in laser-melted ductile iron depended on the solidification rate. Dendritic retained austenite with a continuous interdendritic carbide, with a microhardness from 400 to 650 HV, formed at high solidification rates. A lamellar mixture of ferrite and cementite plates, with a microhardness from 1000 to 1250 HV, formed at low solidification rates.

Bamberger *et al.* (Ref 24) found that in laser-melted gray cast iron, full dissolution of the graphite occurred, leading to the formation of hot tears on the surface. On the other hand, in laser-melted nodular cast iron, partial dissolution of the graphite occurred, resulting in increased ductility and no hot tearing. Steen (Ref 21) found that martensitic structures were formed in some stainless steels, whereas fine carbide dispersions were formed in tool steels.

In laser-melted titanium alloys, highly dislocated, fine, plate-like martensitic structures were reported by Steen (Ref 21). In Ti-15Mo, the crystal growth was epitaxial from the β grains in the substrate in a direction normal to the interface (Ref 25). In Ti-6Al-4V, β martensite transformation in the solid state occurred upon rapid cooling. Plates of a twinned substructure were common. In Ti-10V-2Fe-3Al, martensite with some α formed upon solidification.

Wear Behavior of Laser-Melted Surfaces. Bergmann (Ref 22) has summarized the wear properties of laser-melted cast irons. For laser-melted gray iron containing flake graphite in a pearlitic matrix, the dry pin-on-disk test demonstrated that the wear behavior improved by an order of magnitude and was better than a fully martensitic structure. In a test involving rolls that ran against one another with a fixed relative slip, it was found that the wear resistance of laser-melted ductile iron was superior to that of: 0.6% C steel, case-hardened 16MnCr5 steel, nitrided or carburized 16MnCr5 steel, and a gas tungsten arc welded melted surface.

Ju *et al.* (Ref 26) reported significant improvement in erosive wear behavior in laser-melted ductile and gray irons, as shown in Fig. 7. This improvement was due to the presence of mechanically metastable austenite that transforms to martensite when plastic deformation of the near-surface region occurs during wear.

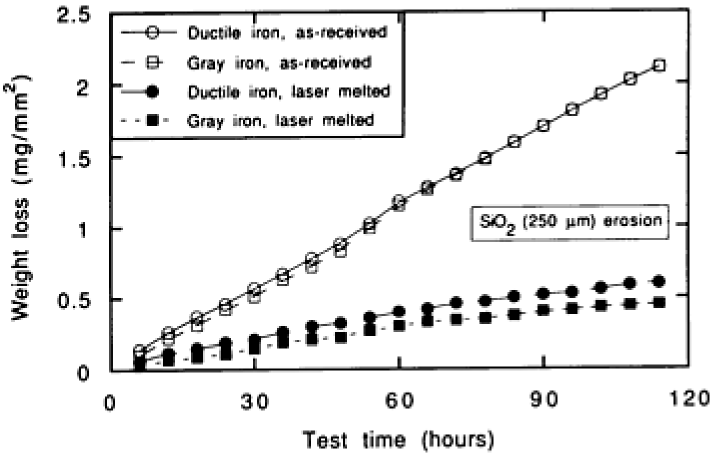


Fig. 7 Erosive wear behavior of as-received and laser-melted gray and ductile irons. Source: Ref 26

Tomlinson *et al.* (Ref 27) reported that laser melting reduced the amount of cavitation erosion of gray cast iron in distilled water by a factor of 0.3, and, in 3% saltwater, by a factor of 0.57.

Kusinski and Thomas (Ref 28) found that by laser melting, the wear resistance of Fe-Cr-Mn-C steels was improved by a factor of 1.4 to 1.6 over that of conventionally hardened steels only when the melt zone was shallow. This improvement was due to the formation of a stress-induced martensite with a highly dislocated lath and internally twinned structure.

De Beurs and De Hosson (Ref 29) found stress-induced hardening in laser-melted chromium steel. During wear, the austenite structure with segregated M_3C carbides, which had a microhardness of 550 HV, transformed to αmartensite, which had a microhardness of 900 HV. In laser-melted steels, the improvement in wear resistance was best when the dendritic growth of the austenite was perpendicular to the surface, because in this orientation the concentration of carbides was the highest (Ref 30).

Hsu and Molian (Ref 31) reported that the tool life of laser-melted M2 steel tool bits was from 200 to 500% higher than if they were conventionally hardened using catastrophic failure criterion (Fig. 8a). For laser-melted M35 steel tool bits, the tool life was from 20 to 125% higher than if they were conventionally hardened using flank wear failure criterion (Fig. 8b). High-alloy martensite, fine austenite grain size, and finely dispersed carbides all contributed to high hardness, good toughness, and low coefficient of friction.

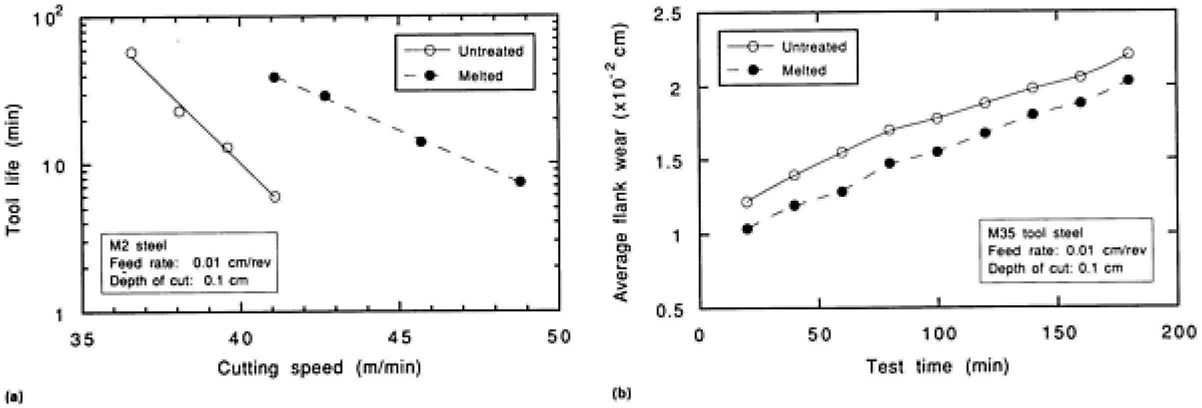


Fig. 8 Tool life of conventionally heat-treated and laser-melted tool bits. (a) M2 tool steel. (b) M35 tool steel. Source: Ref 31

Coquerelle and Fachinetti (Ref 32) laser melted Al-Si alloys and found that the resulting reduction in dendrite size and Si particle size improved wear resistance. Using a pin-on-disk wear measurement technique, the wear volume was reduced 10-fold for a hypoeutectic alloy. The wear rate decreased with increasing silicon content as long as the silicon particle size was less than 2 μm (80 μin.).

Applications of Laser Melting. The benefits of laser melting cast irons are the achievement of very high hardness and wear resistance on a relatively cheap material that is tough in the bulk, in the case of ductile iron, and easy to shape. Tool steels can also be laser surface melted to give very high hardness and wear resistance properties. Although several studies on specific alloys have been performed, there are very few commercial applications for laser surface melting to improve wear resistance. However, during laser transformation hardening, a very thin surface layer is sometimes melted, and the tribological benefits of laser melting are realized in the laser-hardening examples.

Bergmann (Ref 22) describes the application of laser melting to a cast iron rocker arm and camshaft. The laser-melted components achieved a lifetime that was four times better than was achieved by melting using gas tungsten are welding. A German automotive manufacturer routinely uses laser melting to harden camshafts.

Laser Alloying

Processing. A technique of localized alloy formation is laser surface melting with the simultaneous, controlled addition of alloying elements. These alloying elements diffuse rapidly into the melt pool, and the desired depth of alloying can be obtained in a short period of time. By this means, a desired alloy chemistry and microstructure can be generated on the sample surface; the degree of microstructural refinement will depend on the solidification rate. The surface of a low-cost alloy, such as mild steel, can be selectively alloyed to enhance properties, such as resistance to wear, in such a way that only the locally modified surface possesses properties typical of tribological alloys. This results in substantial cost savings, and reduces the dependence on strategic materials. Typical processing parameters for laser alloying are a power density from 10 to 3000 MW/m^2 (6.5 to 1935 kW/in.^{2}) and an interaction time from 0.01 to 1 s. An inert shielding gas is normally used.

One method of alloying is to apply appropriate mixtures of powders on the sample surface, either by spraying the powder mixture suspended in alcohol to form a loosely packed coating, or by coating a slurry suspended in organic binders (Ref 33). The use of metal powders in laser alloying is the least expensive, but, with appropriate process modifications, alloys in the form of rods, wires, ribbons, and sheets can also be added. Because of inconsistency in coating application and possible loss during processing, the composition of the surface alloy may not reflect that of the applied alloying elements.

Powders that are added in controlled quantities using powder feeders with electronic metering can reduce this problem. Powders usually range from 10 to 50 μm (0.4 to 2 mils); the finer size facilities dissolution and alloying. Typically, the powder flow rate is from 0.005 to 0.1 cm^3/s (0.0003 to 0.006 in. ^{3}/s), particle velocity is from 1 to 5 m/s (3.3 to 16.4 ft/s), and carrier gas velocity is from 3 to 10 m/s (10 to 33 ft/s). Whether preplaced or fed, the powder increases the coupling coefficient of the surface, thereby avoiding the need for special absorptive coatings.

During surface alloying, temperature gradients form in the melt pool, which, along with the addition of alloying elements, influences the surface tension. Convection currents are established as a result of the surface tension gradients, and can cause variations in concentration. If the laser beam is oscillated, the melt spreads out over a wider region and, because the beam sweeps the same area several times, a potentially beneficial mixing action can occur. Any alloy concentration gradients that may have formed initially are thus diminished. If the melt pass is made with a very rapid sample translation rate that is greater than 50 mm/s (2 in./s), then inhomogeneities in the microstructure can result, but the high quench rate ensure minimal segregation. The rapid quench also facilitates alloying with hard-to-alloy elements, such as iron, chromium, carbon, and manganese. Examples of laser alloying are cast irons with chromium, silicon, and carbon; steels with chromium, molybdenum, boron, and nickel; stainless steels with carbon; aluminum with silicon, carbon, nitrogen, and iron; and superalloys with chromium.

Another method of laser alloying is gas reaction, for which a shielding gas of appropriate composition is chosen (Ref 34). For the nitriding of titanium alloys, a dilute (10 to 20%) mixture of N_2 in argon is used as the reactive gas. The extent of nitriding depends on the partial pressure of N_2 in the atmosphere, which is determined by the law of mass action, and the temperature, which depends on the laser power density. Alloys of titanium can be carburized by using a methane/argon mixture, and carbonitrided by using a methane/N_2 mixture.

Folkes *et al.* (Ref 25) nitrided titanium alloys by blowing N_2 with a flow rate of 100 to 250 cm^3/s (6 to 15 $in.^3/s$) onto the melted surface. Influenced by the mixing effects derived from surface tension forces and from the impingement of the N_2 jet on the surface, nitrogen was uniformly distributed and TiN formed throughout the melt zone. The authors also carbonitrided titanium alloys by applying a carbon-base paint over the surface and then laser melting in the presence of a N_2 gas jet.

Examples of Laser-Alloyed Surface Microstructures. With laser alloying, a microstructure that strengthens the surface is produced to develop resistance to wear. Surface alloying of AISI 1018 steel with carbon and chromium produces stable carbides, such as M_7C_3 and M_3C, in austenitic, pearlitic, or martensitic matrices (Ref 33). The microhardness of these carbides is from 1100 to 1200 HK, and, when uniformly dispersed in a pearlitic or martensitic matrix, results in an improved resistance to abrasive and adhesive wear. With a 16% addition of chromium, the microstructure of laser-alloyed 1018 steel is martensitic, with small islands of ferrite, as shown in Fig. 9(a). With additions of 43% Cr and 4.4% C, the microstructure consists of hexagonal-shaped M_7C_3 carbides, as shown in Fig. 9(b).

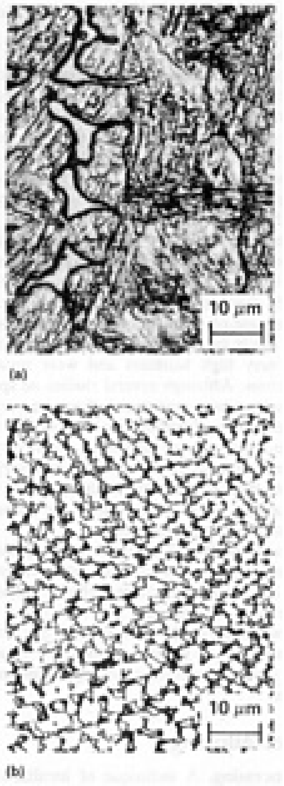


Fig. 9 Microstructures of laser-alloyed 1018 steel. (a) Addition of 16% Cr. (b) Addition of 43% Cr and 4.4% C. Source: Ref 33

Marsden *et al.* (Ref 35) found that laser melting of graphite-coated stainless steels produced dendritic austenite and interdendritic carbides with a microhardness of 600 HV, whereas carbon-powder-injected microstructures contained martensitic cells, with segregated carbon and chromium forming a small amount of eutectic in the intercellular regions. With increasing carbon content, the hardness was increased to 1500 HV, because of the increase in carbide volume.

The reactivity of titanium makes its alloys ideal targets for laser alloying by gas reaction. TiC, TiN, and carbonitrides form as dendrites in a titanium matrix throughout the melted area. The matrix is α' in Ti-6Al-4V (Ref 34) and β in Ti-15Mo (Ref 25). Laser gas alloying can also be used to increase the carbon content of low-carbon steels, and to produce ledeburite structures on steels.

Wear Behavior of Laser-Alloyed Surfaces. Yu and Quan (Ref 36) laser alloyed steel with carbon, nitrogen, and boron and found that the abrasive wear resistance improved 13 times, compared to that of unalloyed steel. This improvement was due to the hardening caused by a uniform dispersion of nitrides, carbides, and borides in a fine dendritic ferritic matrix. Rieker *et al.* (Ref 37) laser alloyed a free-machining ferritic stainless steel with various combinations of

molybdenum and boron by applying the elements as paint. For chemical and structural homogeneity, a second remelting was done. As shown in Fig. 10, the wear behavior improved under all alloying conditions. A microcrystalline ferritic structure with a dispersion of boride particles was responsible for the improved abrasive wear resistance.

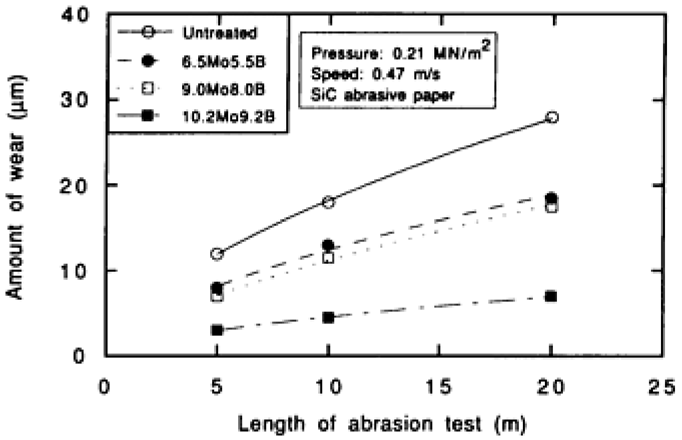


Fig. 10 Wear behavior of untreated and laser-alloyed ferritic stainless steel. Source: Ref 37

Using a dry-sand-rubber wheel abrasion test, Tan and Doong (Ref 38) found that when AISI 1018 steel was layer alloyed with chromium, the abrasion resistance improved because of the formation of a martensite-ferrite structure. Xiaoping *et al.* (Ref 39) laser alloyed low-carbon steel with two Co-Cr-W combinations. With 17Co7Cr3W (815 HV), abrasive wear resistance was better, whereas with 52Co22Cr9W (625 HV), corrosive wear resistance was better. In the former, martensite formed, whereas in the latter, a mixture of γ-Co and $Cr_{23}C_6$ formed.

Huntz *et al.* (Ref 40) improved the pin-on-disk wear behavior of 70Ni30Cr by laser melting the alloy that was precoated with carbon. The formation of a variety of chromium-rich carbides in a γ-phase solid solution contributed to the improvement in wear resistance.

Mordike (Ref 34) reported improved rolling and adhesive wear resistance of titanium by laser gas alloying. The increased surface hardness improved resistance to penetration and, thus, the resistance to plowing wear and erosion.

Applications of Laser Alloying. Laser alloying has been primarily applied to improved corrosion resistance. A very common technique is to alloy steels with chromium. The only examples of laser alloying to improve wear resistance are exhaust valves fabricated by Fiat Research Laboratory.

Laser Cladding

Processing. In laser cladding, a necessarily different material is fused to the substrate surface with minimal dilution. The cladding has a microstructure and properties that are entirely different from those of the substrate. Lasers enable the cladding of high-melting-point alloys over low-melting-point workpieces (Ref 33). Cobalt-, nickel-, and iron-base hardfacing alloys are the usual cladding materials used for metal-to-metal contact, impact, erosion, and abrasion wear resistance.

Cladding material selection depends on its metallurgical compatibility with the substrate material. Alloys that form brittle intermetallic phases with the substrate are undesirable. For example, cobalt forms harmful intermetallics with titanium, and cobalt-base cladding alloys would be unsuitable on titanium alloy substrates.

The cladding alloy is melted by a laser beam and allowed to spread freely and freeze over the substrate. The beam also melts a very thin layer of the substrate, which combines with the liquid cladding to the least extent necessary and solidifies to form a strong metallurgical bond. A good fusion bond can be achieved with a dilution zone that is only from 10 to 20 μm (0.4 to 0.8 mil) thick.

The clad alloy can be in several forms, examples of which are a prealloyed powder that is applied to the sample surface with or without a binder, a self-fluxing powder that is flame sprayed, a hard-facing alloy that is plasma sprayed, or a chip that is preplaced. Laser consolidation of these coatings results in densification and smoothening, eliminates channels to the substrate, improves the bonding between coating and substrate, and reduces porosity, all of which contribute to the strength and integrity of the clad layer. Processing parameters for laser cladding are a power density that ranges from 10 to 1000 MW/m^2 (6.5 to 645 $kW/in.^2$) and an interaction time from 0.1 to 1 s. The shielding gas could be any of the inert gases or a combination of gases, such as He/Ar and H_2/Ar.

It is sometimes difficult to keep preplaced powder in place as it is melted by the laser beam. The external addition of powder avoids this problem. The clad material can be gravity-fed just ahead of the beam spot, so that as the sample translates, the laser beam melts and fuses the powder to the workpiece.

With the help of a suitable carrier gas, the powder can also be blown into the laser-melted pool. In this case, the melt pool is mostly clad material, with the leading edge incorporating a small portion of the substrate surface. Particle size usually ranges from 45 to 150 μm (1.8 to 6 mils), with the larger size helping to minimize alloying.

The amount of dilution and clad thickness are controlled by the interaction time and the powder flow flux. Desirable dilutions are from 3 to 8%. Clad thickness can be varied from 0.05 to 4 mm (2 to 155 mils). The optimum powder flow flux is one that gives minimum dilution and maximum deposition (Ref 1). The lowest particle velocity needed to uniformly convey the powder to the melt pool is desirable to minimize losses from the molten pool. The powder delivery nozzle is positioned in such a manner as to flood the entire melt pool with powder. The feeding angle is generally from 35 to 45° to the horizontal, and the feed tube, which typically has a 3 mm (0.12 in.) diam, is positioned from 10 to 12 mm (0.4 to 0.5 in.) from the substrate. Typically, the powder flow rate is from 0.005 to 0.1 cm^3/s (0.0003 to 0.006 $in.^3/s$), particle velocity is from 1 to 2 m/s (3.3 to 6.6 ft/s), and carrier gas velocity is from 3 to 7 m/s (10 to 23 ft/s). Uniform feeding of the powder ensures uniform clad layers. Unlike plasma-sprayed coatings, porosity and unmelted powder particles are almost never observed within a laser-clad layer.

The shape of the cladding can be controlled within certain limits, the most preferred shape being one with an obtuse angle to the substrate. Multiple clad tracks produced by overlapping such a clad layer will have a good surface finish and be porosity-free. Cladding is commonly achieved by relatively large "doughnut-shaped" beams, or by somewhat focused beams that are oscillated. These ensure a desirable clad profile. Clad material can also be added in rod, wire, or sheet form, but special procedures are needed because of reflectivity problems.

Thermal stresses in the cladding can cause harmful cracking, but this can be eliminated by an appropriate preheating practice. Low power densities, large beam diameters, and slow sample translation rates tend to produce crack-free clad layers.

Ayers and Gnanamuthu (Ref 33) give several examples of laser cladding for improved wear resistance, such as Tribaloy T-800 on powder-coated ASTM A 387 steel; Haynes Stellite alloy No. 1 cast rods placed on AISI 4815 steel; and a mixture of WC and Fe powder applied without binder on AISI 1018 steel.

In addition to stainless steel on mild steel, Steen (Ref 1) has reported successful laser cladding of nickel, bronze, Fe-B, and Stellite SF6 on mild steel; chromium on titanium; stainless on aluminum; mild steel on stainless, and stellite on brass. Other examples of laser cladding are AISI 1016 steel clad with a mixture of chromium, manganese, and carbon powders (Ref 41) and gray cast iron clad with a powder-fed mixture of molybdenum, Cr_2C_3, chromium, nickel, and silicon (Ref 42).

Examples of Laser-Clad Layer Microstructures. The key to improved wear resistance is the presence of hard phase distributions in the laser-clad microstructures. The wear resistance of Tribaloy is attributed to the presence of hard Laves intermetallic phases surrounded by solid solution and a eutectic matrix (Ref 33). In T-800 alloy, the Laves phases make up nearly 50 vol% of the microstructure (Fig. 11a), and the cladding has an average microhardness of 690 HK. Stellite alloy No. 1 contains M_7C_3 and M_6C carbides finely dispersed in an austenitic cobalt matrix (Fig. 11b), and has a microhardness of 730 HV. Microstructural analysis of the WC-Fe mixture revealed a distinct region on the WC particles, which formed because of interaction with the liquid iron. Microhardness of the WC particles was 1100 HK, and that of the surrounding matrix was 870 HK.

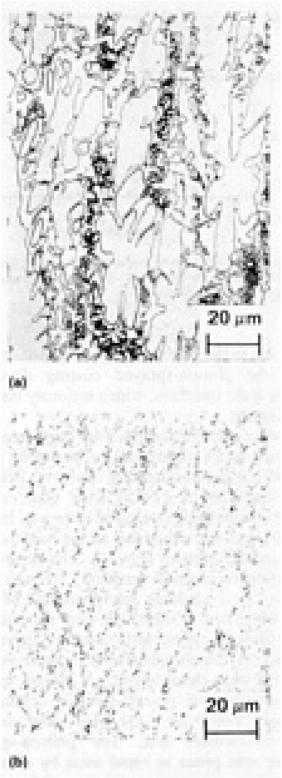


Fig. 11 Examples of laser-clad microstructures. (a) Tribaloy T-800 alloy on ASTM A 387 steel. (b) Haynes Stellite alloy No. 1 on AISI 4815 steel. Source: Ref 33

The Fe-Cr-Mn-C cladding of Mazumder and Singh (Ref 41) had a microstructural consisting of M_7C_3 and M_6C type carbides in a ferritic matrix, a solid solubility extension of chromium in ferrite by 50%, and a microhardness of 550 HV. The molybdenum-clad coating on gray iron of Belmondo and Castagna (Ref 42) consisted of molybdenum dendrites surrounded by a Cr-Ni matrix containing Cr_2C_3, which had a microhardness of 900 HV.

Wear Behavior of Laser-Clad Layers. Mazumder and Singh (Ref 41) found that the Fe-Cr-Mn-C cladding on AISI 1016 steel resulted in tribological properties that were superior to Stellite 6. The width of the wear scar was reduced from 3.5 mm (0.14 in.) in the base alloy to about 0.6 mm (0.02 in.) in the laser cladding. Under the same test conditions, Stellite 6, a common hardfacing alloy, developed a wear scar which of 1.4 mm (0.06 in.).

Abbas *et al.* (Ref 43) laser clad a mild steel with Stellite 6, Alloy 4815, and their composites with SiC by using pneumatic powder delivery. Wear tests, conducted by grinding the samples against a revolving alumina disk, showed that the composite clad samples had the best abrasive wear resistance (Fig. 12).

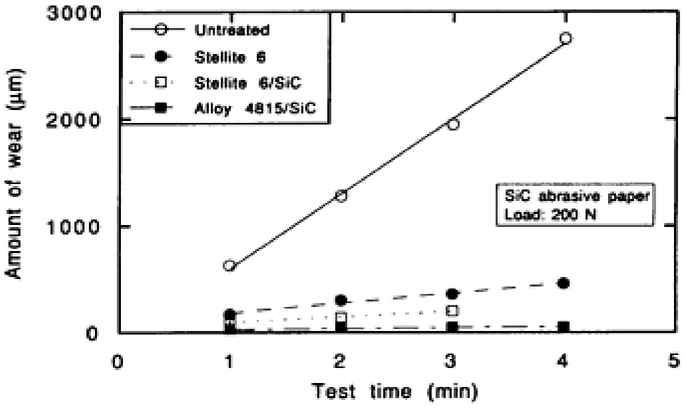


Fig. 12 Abrasive wear behavior of mild steel laser clad with Stellite 6, Stellite 6/SiC, and Alloy 4815/SiC. Source: Ref 43

Belmondo and Castagna (Ref 42) performed wear tests on the Mo-Cr_2C_3 clad cast iron samples using a reciprocating motion testing machine. When compared with plasma-sprayed coatings of similar composition, the performance of laser-clad coatings was far superior under all testing conditions, especially at the highest pressures. With a load of 25 MPa (3.6 ksi) and slider roughness of 0.3 μm (12 μin.), the wear loss in the laser sample was 0.4 mg (1.4×10^{-5} oz), compared to 1.3 mg (4.6×10^{-5} oz) in the plasma-sprayed sample. In addition, the plasma-sprayed coating developed cracks at the interface, which seriously impaired its integrity.

Molian and Hualun (Ref 44) performed pin-on-block reciprocating sliding wear tests on Ti-6Al-4V alloy substrate laser clad with BN, both with and without a NiCrCoA1Y addition, and found that the wear resistance improved from 10 to 200 times more than that of age-hardened and laser-melted samples (Fig. 13). The presence of solidification reaction products, TiN and TiB_2, resulted in a cladding microhardness of 1600 HV, which led to improved wear resistance by preventing ploughing and reducing friction.

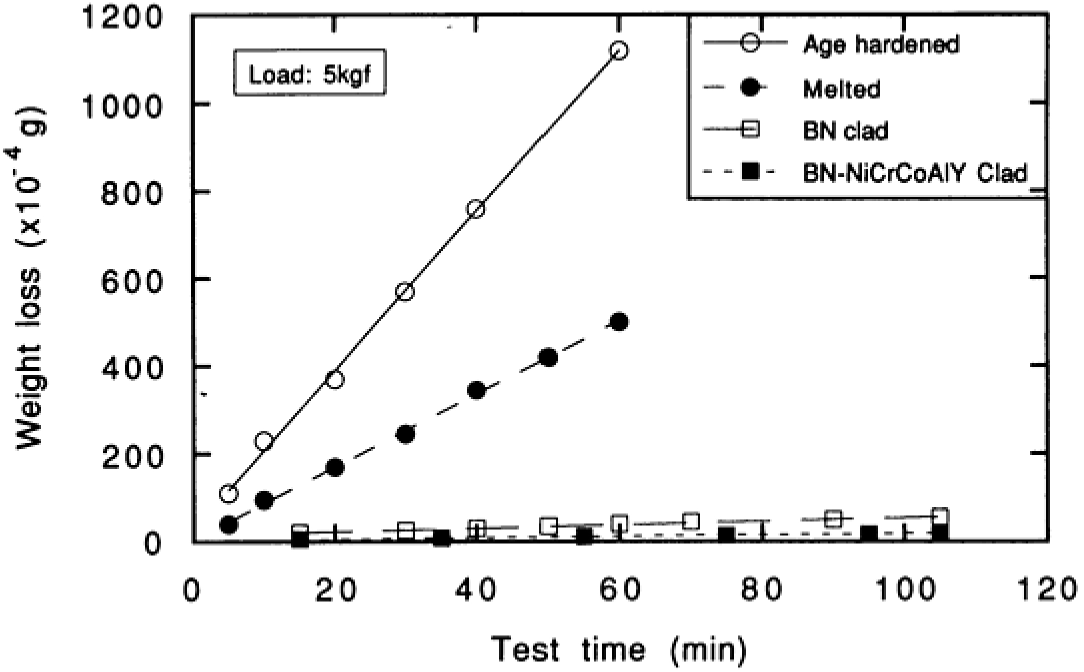


Fig. 13 Sliding wear behavior of laser-clad Ti-6Al-4V alloy. Source: Ref 44

Boas and Bamberger (Ref 45) determined the abrasive wear characteristics of plasma-sprayed and laser-melted plasma-sprayed coatings of Tribaloy T-400 on AISI 4130 steel using a block-on-cylinder test. The plasma-sprayed coating was prone to rapid wear by delamination, whereas laser consolidation of the plasma coating removed flaws, such as porosity and microfissures, and generated an adherent wear-resistant layer. Figure 14 demonstrates the improvement in wear after laser consolidation of the plasma coating. For comparison, results on a hard D2 steel are also given.

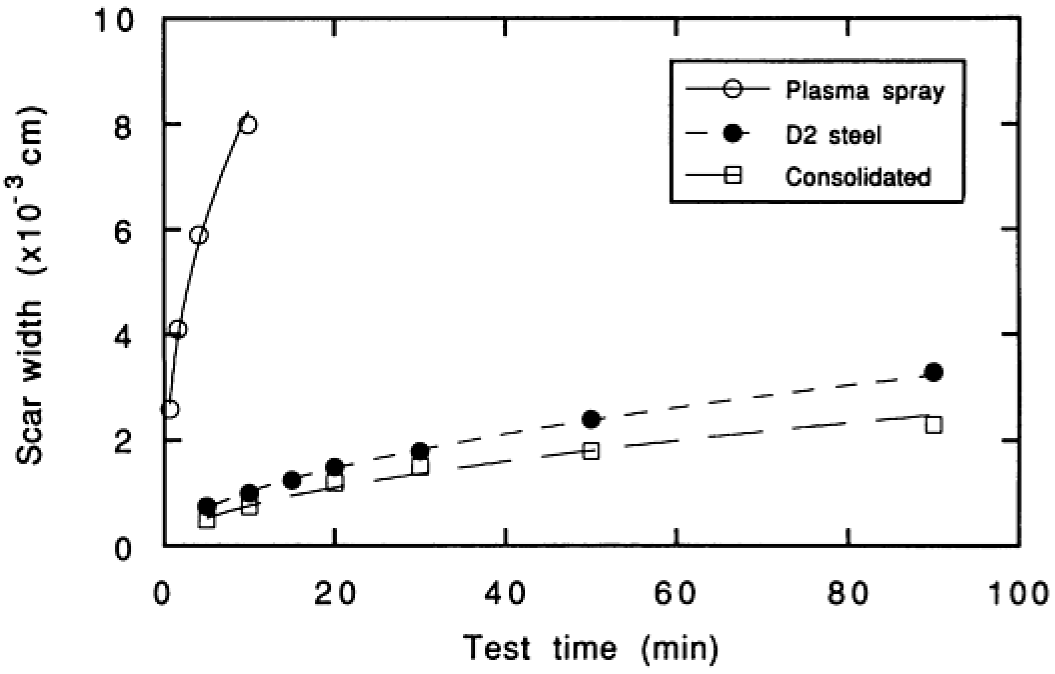


Fig. 14 Wear scar growth curves of plasma-coated 4130 steel before and after laser consolidation. Source: Ref 45

Laser-Cladding Applications. Because one of the primary goals of laser cladding is to improve the tribological properties of components, several applications have been found for the technique. Some of them are in the exploratory stage, others are in the pilot-plant stage, and a few have reached the production stage.

Bruck (Ref 46) demonstrated cladding within a confined space by coating the inside surface of small-bore pipes (inside diameter of 50 to 100 mm, or 2 to 4 in., and length of 0.3 to 1.2 m, or 1 to 4 ft), as shown in Fig. 15. An oscillating mirror placed in the pipe formed a 12.5 mm (0.5 in.) wide melt pass, into which the clad alloy powder was fed. The pipe was rotated and translated to cover the entire inner surface. Previously, chrome plating was used, but the laser technique appreciably improved galling resistance. Using suitable cooling techniques, the pipe temperature was maintained below 500 °C (930 °F) to avoid deterioration of the core properties, as well as distortion.

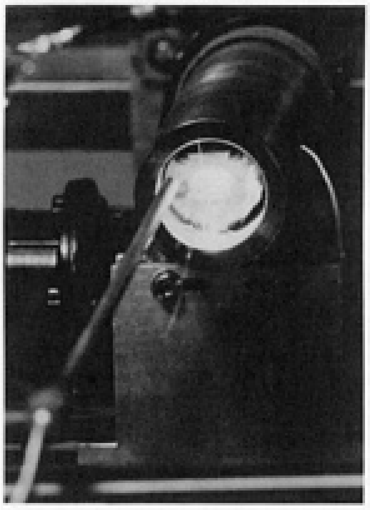


Fig. 15 Laser cladding of small-bore pipes. Source: Ref 46

La Rocca (Ref 19) describes laser cladding of exhaust valves with stellite at an Italian automotive manufacturer. The stellite powder was fed into the laser melt pool. The laser-clad material was superior to gas tungsten arc welded material in terms of thickness uniformity and uniform microstructure and elemental distributions, and had better adhesion. Powder utilization also was better (30% that of gas tungsten arc welding), and over-metal removal was reduced by 10 to 15%. One British automotive manufacturer uses production-stage laser cladding of a nickel-alloy turbine blade shroud interlock by powder feeding either Tribaloy or Nimonics (Ref 47). Previously, a manual arc melting technique was used. Laser cladding reduced cladding time from 14 min to 75 s, improved productivity and quality, reduced cost by 85%, and reduced powder consumption by 50%. Table 2 lists components that are laser clad for commercial applications.

Table 2 Component, cladding alloy, and cladding technique

Component	Clad alloy	Cladding technique	Company	Ref
Gate and seat of steel valves for oil-field, geothermal, and nuclear energy production, and chemical processing	Stellite 6	Powder feed	W-K-M Div. of Joy Mfg., Houston, TX	48
Diesel engine valve	Stellite SF6	Preplaced powder paste	Cabot Corp., Kokomo, IN	49
Gate valve seat	Deloro 60	Preplaced powder paste	Cabot Corp., Kokomo, IN	49
Gas turbine blade "Z" notch	Co-base alloy	Powder feed	Quantum Laser, Edison, NJ	50
Leading edge of turbine blades	Co-base alloy	Powder feed	Quantum Laser, Edison, NJ	50
Pump bushings and impellers	Co-base alloy	Powder feed	Quantum Laser, Edison, NJ	51
Engine valve bases	Ni-Cr alloy	Powder feed	Quantum Laser, Edison, NJ	51
Oil-field valve gates	WC and Co alloy	Powder feed	Quantum Laser, Edison, NJ	51
Tractor bushings	Stellite 6	Powder feed	Quantum Laser, Edison, NJ	51
Turbine blades, pump valve seats	Stellite, Colmoloy, blended powder	Preplaced beds, gravity feed	Westinghouse, Pittsburgh, PA	52
Valve stem, valve seat, aluminum block	CrC_2, Cr, Ni, Mo/cast Fe	Preplaced powder	Fiat, Turin, Italy	52
Off-shore drilling and production parts, valve components, boiler firewall	Stellite, Colmonoy, alloys/carbides	Powder feed	Combustion Eng., OH	52
Aerospace components	Stellite, Tribaloy	Powder feed	Rockwell Int., CA	52

Turbine blade shroud interlocks	PWA 694, Nimonics	Preplaced chip	Pratt & Whitney, West Palm Beach, FL	52
Turbine blade "Z" notch	Hardfacing	Powder feed	United Technol., E. Hartford, CT	53
Swivel joints, petroleum valves	Hardfacing	Powder feed	Spectra Physics, San Jose, CA	53
Gas engine turbine blades	Hardfacing	Powder feed	General Electric, SC	54

Laser Melt/Particle Injection

Processing. Laser melt/particle injection produces an *in situ,* metal-matrix/particulate composite surface layer by mixing, but not melting, the second phase with the substrate. The particulate material is injected with sufficient velocity as a spray into the melt pool formed by the laser beam. If the second phase is hard, such as a carbide, the injected layer can be made to resist wear.

Figure 16 depicts the laser injection process. An oscillating CW CO_2 laser beam produces a shallow melt pool on a substrate, into which are injected powder particles via a nearby nozzle. As the sample translates, an injected layer forms on the surface. By varying the amplitude of oscillation, melt widths ranging from 3 to 20 mm (0.12 to 0.8 in.) are possible (Ref 55). Processing conditions for laser melt/particle injection are a power density from 10 to 3000 MW/m^2 (6.45 to 1935 $kW/in.^2$) and an interaction time from 0.1 to 1 s. Any inert shielding gas is normally used.

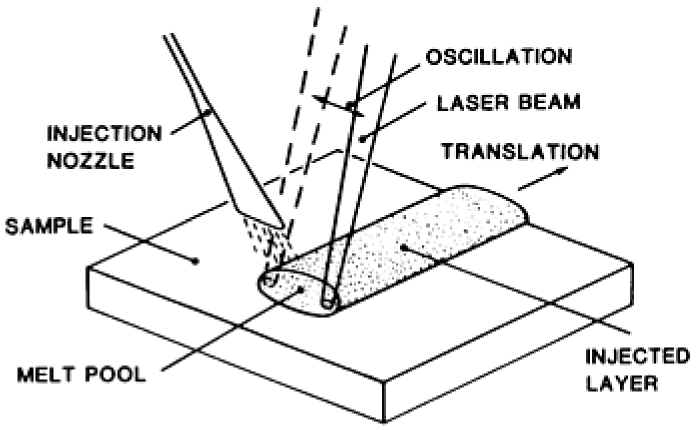


Fig. 16 Laser melt/particle injection process

The particles are propelled by pressurized helium gas. The pressure depends on the particle size and the relative densities of the powder and the molten substrate. Larger and heavier particles require lower gas pressures. Typically, powder flow rates are from 0.1 to 0.5 cm^3/s (0.006 to 0.03 $in.^3/s$) and particle velocities are a few m/s. Carrier gas pressures typically range from 50 to 120 kPa (7.3 to 17.4 psi). The carrier gas also serves as a cover gas and keeps the particles relatively cool, preventing them from melting or fusing. Particle sizes range from 45 to 150 μm (1.8 to 6 mils). Finer particles tend to either fuse or dissolve in the melt, whereas coarser particles do not flow easily.

The injection nozzle, usually made of copper, is inclined at 60° to the horizontal and is positioned from 10 to 20 mm (0.4 to 0.8 in.) away from the sample. Its slotted opening is designed to produce a rectangular spray that is of the same size as the melt pool, to ensure that most of the powder is incorporated into the melt.

Although minimum interaction between the carbide phase and the melt is desired, some dissolution does occur. It was found that the degree of carbide dissolution, as well as other physical characteristics of the injected layer, such as penetration depth, mounding, and carbide volume, could be reasonably controlled. Variations in these characteristics, as well as hardness, with laser power and powder feed rate, are determined for Ti-6Al-4V alloy injected with TiC (Fig. 17) (Ref 56). The carbide volume can be varied from 15% to a limit of 60%. For improved wear resistance, higher carbide volumes are desirable, but the conditions that produce increased carbide volume also increase dissolution. An estimation

of the degree of carbide dissolution is given by matrix microhardness. As the trends in Fig. 17 show, matrix microhardness is higher when the carbide volume is greater.

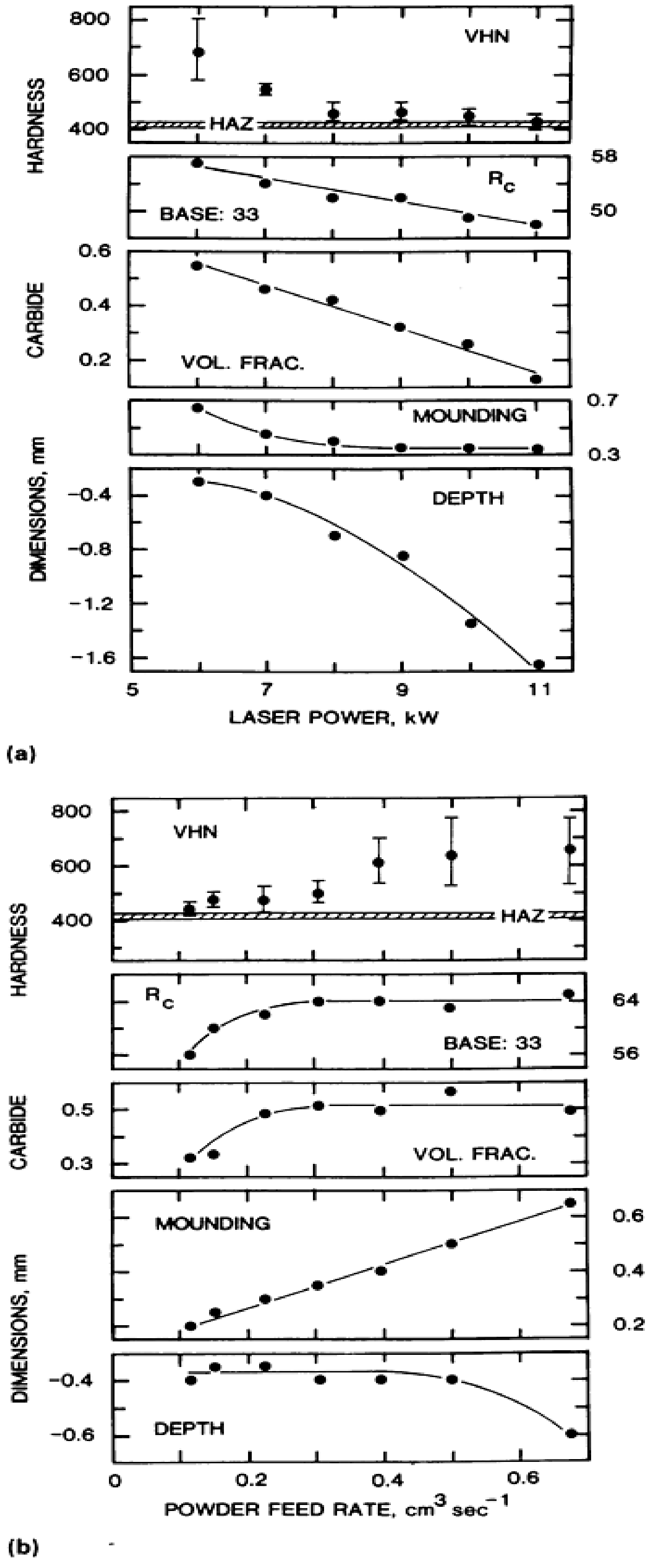


Fig. 17 Effect of processing parameters on injected layer characteristics in Ti-6Al-4V injected with TiC. (a) Laser power. (b) Powder feed rate. Source: Ref 56

Initial work on laser melt/particle injection was done on low-density/high-strength aluminum and titanium alloys, for which few surface-hardening methods exist. Some of the alloy/particulate systems that have been investigated are 5052 Al/TiC, Al bronze/TiC, Ti-6Al-4V/TiC, Ti-6Al-4V/WC, 304 stainless/TiC, 4340 tool/TiC, Inconel 625/TiC, and Inconel 625/WC (Ref 57).

Microstructures of Laser Melt/Particle-Injected Layers. The injected layer is a composite of particles surrounded by a metal matrix. It is desirable that the only surface modification be the presence of the particulate phase, and that the surrounding metal should remain unchanged and retain most of the substrate properties, such as corrosion resistance and toughness. Figure 18 shows the cross-sectional view of a WC injected layer on Inconel 625 alloy substrate. The top surface is rough, but is free of deep channels. The light grains are the WC particles, which are surrounded by the dark Inconel. The carbides are uniformly distributed throughout the injected layer and occupy about 50% of the total volume. Because the matrix phase is essentially made of the same material as the substrate, there is chemical continuity across the interface and a strong metallurgical bond.

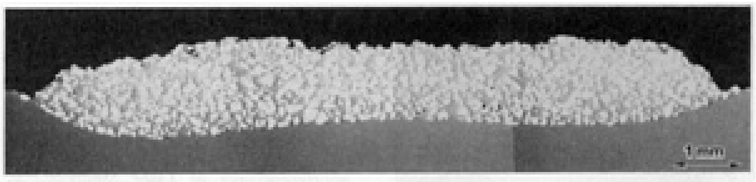


Fig. 18 Cross section of Inconel 625 alloy injected with WC

Carbide dissolution products formed during solidification can influence the matrix microstructure. The carbide particles sometimes develop a perturbed or scalloped, highly alloyed interface with the metal matrix. Eutectic and dendritic carbides are some of the dissolution products that appear within the matrix (Ref 58). Figure 19 shows examples of some of the dissolution products. Besides influencing properties such as microhardness and friction wear, these resolidification products can cause matrix embrittlement and microcracking, although with suitable preheating, cracking can be eliminated (Ref 59).

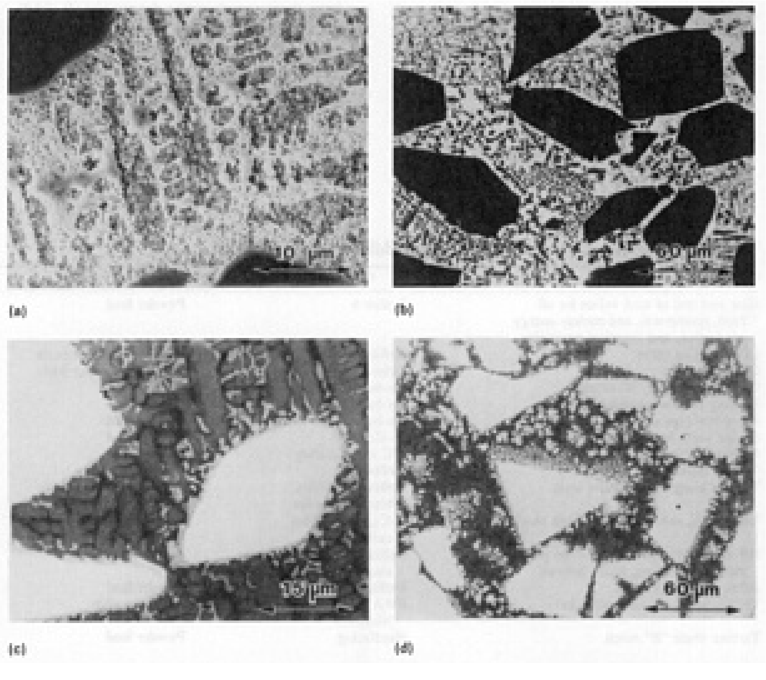


Fig. 19 Carbide morphological changes and matrix resolidification products as a result of carbide phase dissolution in Inconel 625 alloy. (a) Injected with TiC, eutectic carbides. (b) Injected with TiC, dendritic carbides. (c) Injected with WC, eutectic carbides. (d) Injected with WC, dendritic carbides

In order to retain some of the substrate properties in the modified surface, it is important to keep carbide dissolution to a minimum. Some of the matrix microstructural changes are inevitable. Depending on the base alloy, melting and rapid solidification can result in microstructural refinement, formation of solid solutions, phase transformations, and precipitation. These microstructural modifications, along with the dissolution products, can harden the matrix by various mechanisms. Carbide dissolution products harden the matrix by a dispersion-hardening mechanism. This effect was observed in stainless steel, tool steel, Inconel, and titanium alloys. Transformation hardening comes into play when martensite forms in steels and titanium alloys. In tool steels, precipitation hardening also plays a role, and in aluminum bronze, microstructural refinement results in modest hardening (Ref 57).

Wear Behavior of Particle-Injected Surface Layers. On a macroscopic scale, hardening of particle-injected alloys is due to plastic flow inhibition by the injected phase. The relative increase in macrohardness of the composite surface is from 1.1 to 2.6 times, depending on the alloy (Ref 57). On a microscopic scale, hardening of the metal matrix varies with the type of alloy and has been found to affect friction wear behavior (Ref 60).

Friction wear was evaluated using a balt-on-surface test, which measures the coefficient of kinematic friction, μ_k, between a hard steel ball and a polished sample surface. A pressure of 1 N (0.22 lbf) and a slide velocity of 0.0001 m/s (0.02 ft/min) were used. This test is sensitive to minute changes in microstructure. For untreated Inconel 625, μ_k was about 0.7. Particle-injected Inconel 625 had a μ_k that ranged from 0.15 to 0.20 after the first slide, which increased to a range from 0.3 to 0.4 for WC and 0.4 to 0.45 for TiC after several slides (Fig. 20). The wear scar was hardly evident in the optical microscope and there was very little abrasive damage. Although μ_k increased with number of slides, it was not due to continuous wear of the sample surface. Rather, it was due to formation of wear debris.

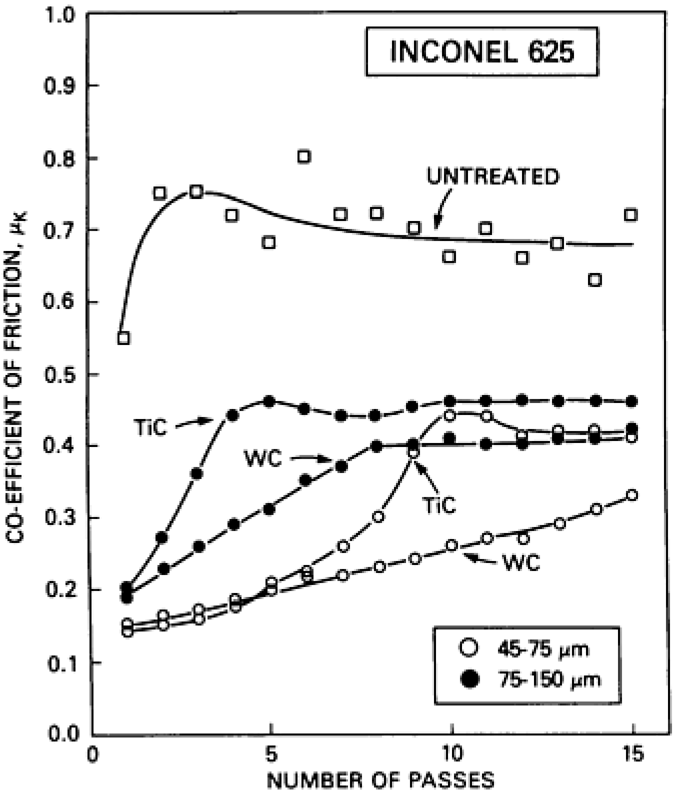


Fig. 20 Coefficient of friction as a function of number of slides in untreated and particle-injected Inconel 625. Source: Ref 60

The wear mechanism was different for the two carbides. With WC injection, the carbide and the harder (600 to 650 HV) metal matrix were strong enough to cause the slider to wear, and the wear debris was mostly that from the steel ball. With TiC injection, the wear debris was a mixture of material from the steel ball and from the softer (400 to 500 HV) metal matrix (Ref 60).

From dry sand/rubber wheel tests done on aluminum- and titanium-base alloys injected with TiC, it was found that the wear volume decreased rapidly with modest volumes of the carbide phase. For aluminum, the reduction is from 0.18 to 0.28 cm^3 (0.01 to 0.02 $in.^3$) to 0.02 to 0.025 cm^3 (0.0012 to 0.0015 $in.^3$). For titanium, it is from 0.06 to 0.065 cm^3 (0.0037 to 0.0040 $in.^3$) to 0.01 cm^3 (0.0006 $in.^3$) (Fig. 21) (Ref 61). Examination of the worn surface showed that the softer metal matrix eroded, but that the hard carbide particles prevented further erosion of the surface. The abrasive action did not uproot the carbides from the matrix, but the normally angular carbides appeared slightly rounded, which helped reduce further erosion.

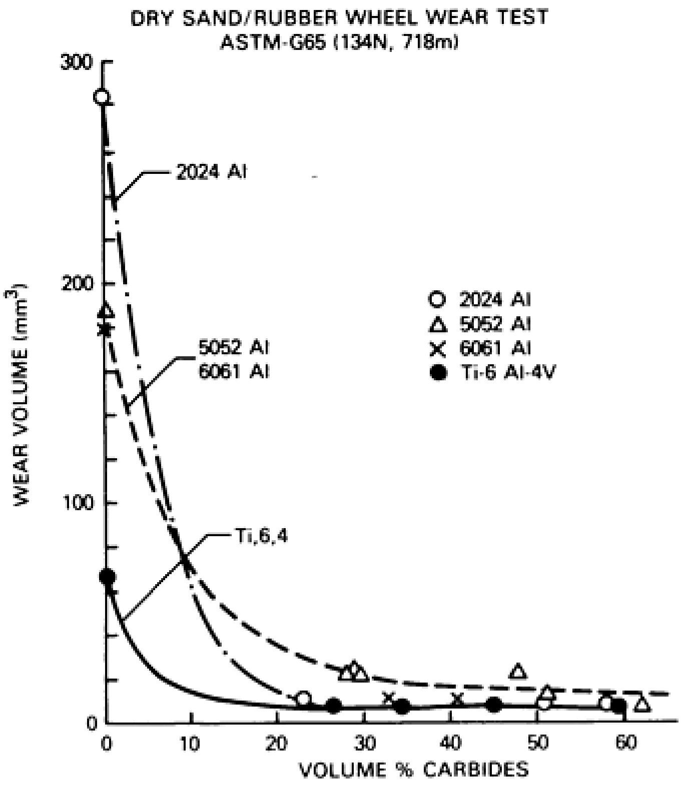


Fig. 21 Reduction in abrasive wear rate with increase in vol% carbide in particle-injected aluminum and titanium alloys. Source: Ref 61

Ayers and Bolster (Ref 62) found that abrasive wear with 3 and 30 μm (120 and 1200 μin.) diamond particles of aluminum and titanium alloy samples was reduced by introducing WC or TiC into their surfaces. The wear resistance of aluminum alloys improved by a factor of 30 in samples containing TiC, whereas for Ti-6Al-4V containing TiC, the improvement was less dramatic (factor of 4).

Applications of Laser Melt/Particle Injection Process. Most of the laser melt/particle injection processing work has involved process optimization, microstructure evaluation, and test sample preparation. One application involved fabrication of wear-resistant surfaces on Inconel alloy shaft seal test rings, such as the one shown in Fig. 22 (Ref 63). The injectant is WC and the circular pass is 127 mm (5 in.) in diameter and 10 mm (0.4 in.) wide. The coverage rate was 130 mm^2/s (0.2 $in.^2/s$). Before testing, the rough surface was ground flat. The seal ring was successfully tested against a graphite mating surface using water as the lubricant.

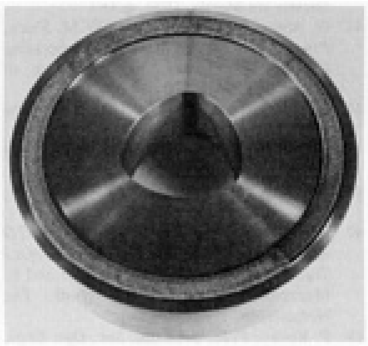


Fig. 22 Inconel alloy shaft seal test ring particle-injected with WC. Source: Ref 63

References

1. W.M. Steen, *Proceedings of the NATO Advanced Study Institute on Laser Surface Treatment of Metals,* C.W. Draper and P. Mazzoldi, Ed., Martinus Njihoff, The Netherlands, 1986, p 369
2. O. Sandven, *Heat Treating,* Vol 4, 9th ed., *Metals Handbook,* American Society for Metals, 1981, p 507
3. F.D. Seaman, *The Industrial Laser, Annual Handbook,* D. Belforte and M. Levitt, Ed., PennWell Books, 1986, p 147
4. J. Taylor, *Metalwork. Prod.,* Sept 1979, p 138
5. J. Meijer, M. Seegers, P.H. Vroegop, and G.J.W. Wes, *Proceedings of the International Conference on Applications of Lasers and Electro-optics,* C. Albright, Ed., IFS Ltd., United Kingdom, 1986, p 229
6. M. Yessik and R.P. Scherer, *Sourcebook on Applications of the Laser in Metalworking,* E. Metzbower, Ed., American Society for Metals, 1981, p 219
7. C. Courtney and W.M. Steen, *Sourcebook on Applications of the Laser in Metalworking,* E. Metzbower, Ed., American Society for Metals, 1981, p 195
8. P.A. Molian, Iowa State University, private communication, 1992
9. S.L. Engle, *Sourcebook on Applications of the Laser in Metalworking,* E. Metzbower, Ed., American Society for Metals, 1981, p 149
10. P.A. Molian and M. Baldwin, *J. Tribol.,* Vol 108, 1986, p 326
11. D.S. Gnanamuthu, *Sourcebook on Applications of the Laser in Metalworking,* E. Metzbower, Ed., American Society for Metals, 1981, p 324
12. G. Lu and H. Zhang, *Wear,* Vol 138, 1990, p 1
13. P.A. Molian and M. Baldwin, *J. Tribol.,* Vol 110, 1988, p 462
14. W.J. Tomlinson, R.F. O'Connor, and T.A. Spedding, *Tribol. Int.,* Vol 21 (No. 6), 1988, p 302
15. D.N.H. Trafford, T. Bell, J.H.P.C. Megaw, and A.S. Bransden, *Heat Treating, Proceedings of the International Conference,* Metallurgical Society, 1983, p 198
16. J.A. Wineman and J.E. Miller, *Sourcebook on Applications of the Laser in Metalworking,* E. Metzbower, Ed., American Society for Metals, 1981, p 209
17. M.J. Yessik, *Optical Eng.,* Vol 17 (No. 3), 1978, p 202
18. F.D. Seaman and D.S. Gnanamuthu, *Sourcebook on Applications of the Laser in Metalworking,* E. Metzbower, Ed., American Society for Metals, 1981, p 179
19. A.V. La Rocca, *Proceedings of the NATO Advanced Study Institute on Laser Surface Treatment of Metals,* C.W. Draper and P. Mazzoldi, Ed., Martinus Nijhoff, The Netherlands, 1986, p 521
20. General Motors, *Sourcebook on Applications of the Laser in Metalworking,* E. Metzbower, Ed., American Society for Metals, 1981, p 227

21. W.M. Steen, *Industrial Laser, Annual Handbook,* D. Belforte and M. Levitt, Ed., PennWell Books, 1986, p 158
22. H.W. Bergmann, *Proceedings of the NATO Advanced Study Institute on Laser Surface Treatment of Metals,* C.W. Draper and P. Mazzoldi, Ed., Martinus Nijhoff, The Netherlands, 1986, p 351
23. C.H. Chen, C.P. Ju, and J.M. Rigsbee, *Mater. Sci. Technol.,* Vol 4 (No. 2), 1988, p 161
24. M. Bamberger, M. Boas, and O. Akin, *Z. Mettallkde.,* Vol 79, 1988, p 806
25. J. Folkes, D.R.F. West, and W.M. Steen, *Proceedings of the NATO Advanced Study Institute on Laser Surface Treatment of Metals,* C.W. Draper and P. Mazzoldi, Ed., Martinus Nijhoff, The Netherlands, 1986, p 451
26. C.P. Ju, C.H. Chen, and J.M. Rigsbee, *Mater. Sci. Technol.,* Vol 4 (No. 2), 1988, p 167
27. W.J. Tomlinson, J.H.P.C. Megaw, and A.S. Bransden, *Wear,* Vol 116, 1987, p 249
28. J. Kusinski and G. Thomas, *Proc. SPIE-Int. Soc. Opt. Eng.,* Vol 668, Society of Photo-optical Instrumentation Engineers, 1986, p 150
29. H. De Beurs and J.Th.M. De Hosson, *Scr. Met.,* Vol 21, 1987, p 627
30. H. De Beurs, G. Minholts, and J.Th.M. De Hosson, *Wear,* Vol 132, 1989, p 59
31. M. Hsu and P.A. Molian, *Wear,* Vol 127, 1988, p 253
32. G. Coquerelle and J.L. Fachinetti, *Proceedings of the 5th International Congress on Applications of Lasers and Electro-optics,* C.M. Banas and G.L. Whitney, Ed., IFS Ltd., United Kingdom, 1987, p 19
33. J.D. Ayers and D.S. Gnanamuthu, *Metals Handbook,* 9th ed., Vol 6, *Welding,* American Society for Metals, 1983, p 793
34. B.L. Mordike, *Proceedings of the NATO Advanced Study Institute on Laser Surface Treatment of Metals,* C.W. Draper and P. Mazzoldi, Ed., Martinus Nijhoff, The Netherlands, 1986, p 389
35. C. Marsden, D.R.F. West, and W.M. Steen, *Proceedings of the NATO Advanced Study Institute on Laser Surface Treatment of Metals,* C.W. Draper and P. Mazzoldi, Ed., Martinus Nijhoff, The Netherlands, 1986, p 461
36. Z.S. Yu and Z.K. Quan, *Mater. Chem. Phys.,* Vol 25, 1990, p 277
37. C. Rieker, D.G. Morris, and J. Steffen, *Mater. Sci. Technol.,* Vol 5, 1989, p 590
38. Y. Tan and J. Doong, *Wear,* Vol 132, 1989, p 9
39. J. Xiaoping, H. Zhuangqi, G. Yunlong, J. Ming, and S. Changxu, *Mater. Res. Soc. Symp. Proc.,* Vol 80, Materials Research Society, 1987, p 331
40. A.M. Huntz, T. Puig, L. Confignal, F. Charpentier, and M. Condat, *Mater. Sci. Eng.,* Vol A121, 1989, p 555
41. J. Mazumder and J. Singh, *Proceedings of the NATO Advanced Study Institute on Laser Surface Treatment of Metals,* C.W. Draper and P. Mazzoldi, Ed., Martinus Nijhoff, The Netherlands, 1986, p 297
42. A. Belmondo and M. Castagna, *Sourcebook on Applications of the Laser in Metal-working,* E. Metzbower, Ed., American Society for Metals, 1981, p 310
43. G. Abbas, D.R.F. West, and W.M. Steen, *Proceedings of the Materials Processing Conference,* Vol 69, Laser Institute of America, 1989, p 116
44. P.A. Molian and L. Hualun, *Wear,* Vol 130, 1989, p 337
45. M. Boas and M. Bamberger, *Wear,* Vol 126, 1988, p 197
46. G.J. Bruck, *J. Met.,* Vol 39 (No. 2), 1987, p 10
47. R.M. Macintyre, *Proceedings of the NATO Advanced Study Institute on Laser Surface Treatment of Metals,* C.W. Draper and P. Mazzoldi, Ed., Martinus Nijhoff, The Netherlands, 1986, p 545
48. P. Koshy, *Proc. SPIE-Int. Soc. Opt. Eng.,* Vol 527, R.R. Jacobs, Ed., Society of Photo-optical Instrumentation Engineers, 1985, p 80
49. S.J. Matthews, *Lasers in Materials Processing,* E. Metzbower, Ed., American Society for Metals, 1983, p 138
50. "Quantum Laser Technical Bulletin No. 1-86," Quantum Laser Corp., 1986
51. B.F. Kuvin, *Weld. Des. Fabr.,* May 1987, p 35

52. G.M. Eboo and A.E. Lindemanis, *Proc. SPIE-Int. Soc. Opt. Eng.,* Vol 527, R.R. Jacobs, Ed., Society of Photo-optical Instrumentation Engineers, 1985, p 86
53. D.A. Belforte, Belforte Associates, private communication, 1987
54. P. Mehta, General Electric, private communication, 1987
55. K.P. Cooper, R. Beigel, and P. Slebodnick, *Proceedings of the 5th International Congress on Applications of Lasers and Electro-optics,* C.M. Banas and G.L. Whitney, Ed., IFS Ltd., United Kingdom, 1987, p 169
56. K.P. Cooper, *International Conference on Surface Modifications and Coatings,* R.D. Sisson, Jr., Ed., American Society for Metals, 1986, p 409
57. K.P. Cooper, *Proc. SPIE-Int. Soc. Opt. Eng.,* Vol 957, G. Sepold, Ed., Society of Photo-optical Instrumentation Engineers, 1988, p 42
58. K.P. Cooper and J.D. Ayers, *Proceedings of the 1988 Conference on Laser Surface Modification,* American Welding Society, 1989, p 115
59. K.P. Cooper and P. Slebodnick, *J. Laser Appl.,* Vol 1 (No. 4), 1989, p 21
60. K.P. Cooper, *J. Vac. Sci. Technol.,* Vol A4 (No. 6), 1986, p 2857
61. J.D. Ayers, *Wear,* Vol 97, 1984, p 249
62. J.D. Ayers and R.N. Bolster, *Wear,* Vol 93, 1984, p 193
63. K.P. Cooper and J.D. Ayers, *Surf. Eng.,* Vol 1 (No. 4), 1985, p 263

Carburizing

C.A. Stickels, Environmental Research Institute of Michigan

Introduction

CARBURIZING has been used since ancient times to improve the wear resistance of steel. A low-carbon steel or wrought iron, when heated in the presence of carbonaceous material to a temperature sufficient to form austenite, will dissolve carbon at the surface When quenched to martensite, the carbon-enriched surface layer dramatically increases in hardness. Metalworking artisans learned to combine carburizing with shaping by hot forging to produce tools that retained much of the toughness characteristic of either wrought iron or low-carbon steel, combined with a hard surface to resist wear.

Carburizing is used today to achieve the same combination of toughness and wear resistance (Ref 1). Power-train gears for helicopters, earth-moving equipment, heavy trucks, and passenger cars are all made of carburized steel. Most roller bearings, straight and tapered, are also carburized. The primary advances in carburizing technology in the last 100 years are the development of high-quality alloy steels designed for carburizing and improved processing methods to allow better control over the composition, microstructure, and properties of the carburized case.

Acknowledgement

The author wishes to thank Dr. David L. Sponseller for reviewing the manuscript.

Carburizing Processes

Gas carburizing is the most-used modern production carburizing technique (Ref 2). Batch or continuous furnaces are used with parts that are either placed on fixtures or loaded in baskets. A carbon-rich furnace atmosphere is provided by an endothermic carrier gas enriched with a hydrocarbon gas, such as methane (natural gas) or propane. Furnaces are often fitted with "sealed oil quenches." Oxidation prior to quenching is prevented by maintaining parts in the carburizing atmosphere during transfer to the quenchant.

Although quenching in oils at temperatures from 50 to 75 °C (120 to 165 °F) is common, many carburized parts are martempered by quenching in oils or molten salts at temperatures from 175 to 200 °C (345 to 390 °F). Because the quench temperature for martempering is above the martensite start temperature of the high-carbon case, the case transforms to martensite during subsequent air cooling.

Other forms of carburizing, described in more detail in Volume 4 of the *ASM Handbook* (1991), include:

- Pack carburizing, where parts are packed in a blend of coke and charcoal with "activators," and then heated in a closed container. This is an old, labor-intensive process, but it is still practiced in tool rooms because facility requirements are minimal.
- Vacuum and plasma carburizing, both of which utilize a vacuum chamber with a partial pressure of hydrocarbon gas as the source of carbon. The gas is ionized in the plasma carburizing process. The prime advantage to these processes is the absence of oxygen in the furnace atmosphere
- Salt bath carburizing, in which the baths contain cyanides, cyanates, or carbon-carbonate blends. These are particularly useful for producing thin carburized cases because the carburizing time can be precisely controlled

Characteristics of Carburized Surfaces

Diffusion of Carbon. Some of the features of carburized cases are due to the fact that they are created by the diffusion of carbon. First, there is a gradual transition in carbon content, as well as a transition in microstructure and mechanical properties, between case and core. As a rule, the deeper the case, the less steep is the slope of the carbon gradient (Fig. 1). The absence of any sharp transition in properties assures excellent adherence of the case.

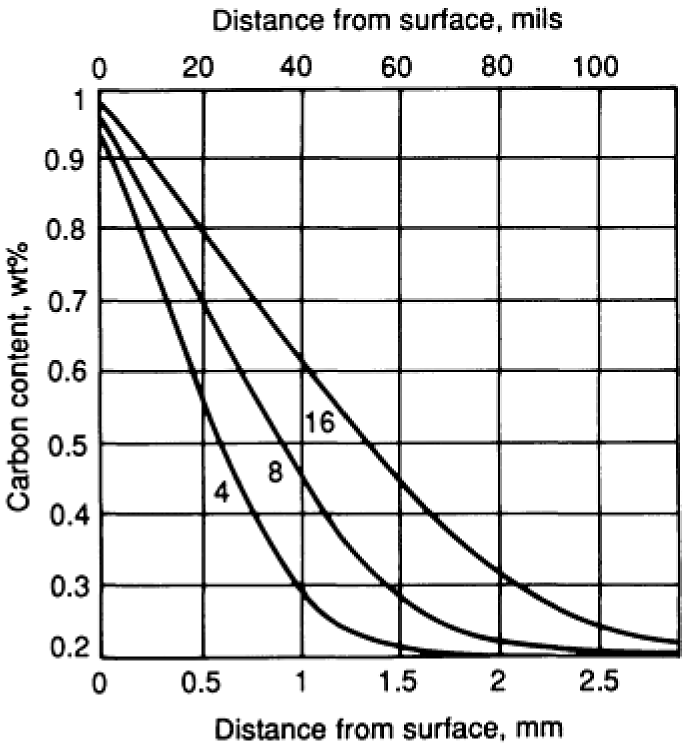


Fig. 1 Computed carbon concentration gradients resulting from gas carburization of SAE 8620 steel for 4, 8, and 16 h at 927 °C (1700 °F). Carbon potential of furnace atmosphere assumed to be 1 wt% C during process

Second, carburized cases are most frequently produced in thicknesses that range from 0.5 to 1.5 mm (20 to 60 mils). At 927 °C (1700 °F), which is a typical processing temperature, this thickness range can be produced in processing times from about 2 to 15 h. Cases as thin as 0.1 mm (4 mils), which require less than 10 min at 927 °C (1700 °F), are sometimes produced on small parts by salt bath carburizing. Cases as deep as 3 mm (120 mils), which require more than two days of carburizing at 927 °C (1700 °F), are occasionally produced on large parts. The ease of adjusting the case depth to resist the anticipated contact loading is one of the great advantages of carburizing.

The microstructure of martensitic steels varies with carbon content in two important ways. First, unalloyed steels with carbon contents of less than about 0.5/0.6 wt% form dislocated lath martensite upon quenching. Steels with carbon contents above 1.0% form twinned plate martensite. Mixtures of lath and plate martensite are found with intermediate carbon levels (Ref 3, 4). Thus, a carburized case will have a mixture of martensite morphologies, with plate martensite dominating in the high-carbon outer layers, and lath martensite dominating toward the core. Alloying elements, which strengthen austenite, promote the formation of plate martensite at lower carbon levels.

Over the same range in carbon as the transition from lath to plate martensite, the amount of austenite retained in the as-quenched structure increases (Fig. 2). However, the amount of austenite retained is a function of other variables, as well. It tends to increase as the quench rate decreases, and as the content of alloying elements increases. Thus, at a given carbon content, the austenite content of as-quenched parts can be higher than the values shown in Fig. 2.

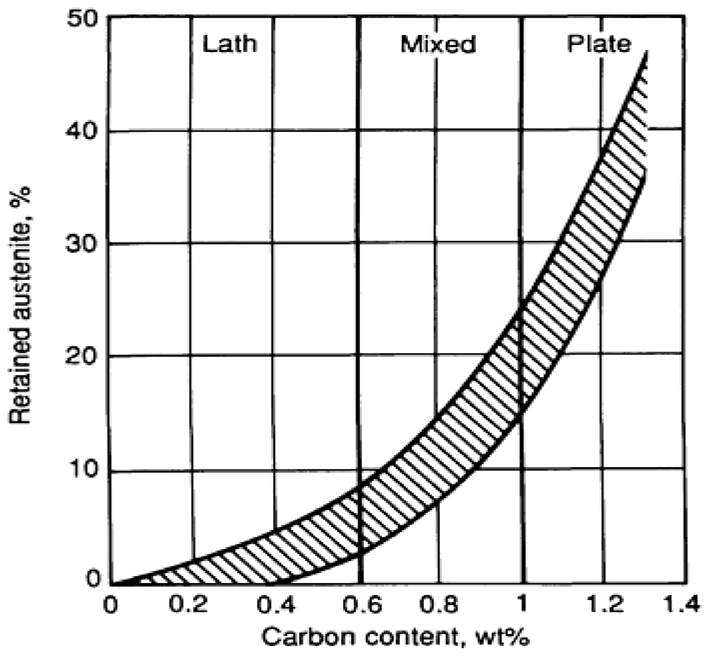


Fig. 2 Percent retained austenite as function of carbon content. Vertical lines show carbon ranges in which lath and plate martensites are found in Fe-C alloys. Source: Ref 5

For example, over 30% retained austenite can be produced in 52100 steel if all the carbon is dissolved during austenitization (Ref 6). In another example, a retained austenite content of 50% has been reported in SAE 8620 steel, with a case carburized to 0.9% carbon (Ref 7).

Although the hardness of martensite increases uniformly with carbon content, the hardness of the martensite-austenite composite obtained on quenching is usually reported to exhibit a plateau above approximately 0.6 wt% carbon (Fig. 3) (Ref 8, 9). In fact, most alloys, when carburized, oil quenched, and tempered at temperatures from 150 to 200 °C (300 to 390 °F), exhibit a maximum in hardness for case carbon contents in the range from 0.6 to 0.9% (Ref 10).

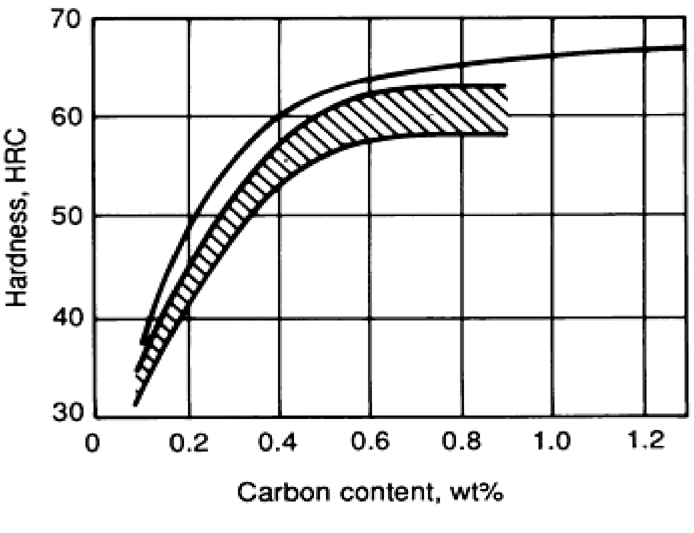


Fig. 3 Rockwell "C" hardness as function of carbon content. Upper curve, after Bain and Paxton (Ref 7), is their estimate of hardness of martensite, with shaded region above 0.9% C representing the uncertainty that is due to retained austenite. Lower band, after Crafts and Lamont (Ref 8), is their estimate of "the maximum hardness usually attained in commercial quenching."

The range in case hardness that is possible at a given carbon level for a variety of carbon steels is much broader than suggested by Fig. 3. (A highly alloyed carburizing steel, such as SAE 9310, may have a case hardness of less than 55 HRC after oil quenching.) When resistance to indentation is the primary concern, the case carbon content is usually specified to achieve the maximum in hardness.

Martensitic structures subjected to alternating shear stresses from repeated contact loading for extended periods of time are found to decompose because of carbon migration aided by dislocation motion (Ref 11). However, microstructural decay does not appear to be a necessary precursor to failure by the usual contact fatigue mechanisms.

Composite structures that contain primary carbides can also be produced by carburizing. Three common morphologies are discussed below.

Coarse primary carbides (from 1 to 10 μm, or 40 to 400 μin.) (Fig. 4) are produced by carburizing in an atmosphere with a carbon potential that is high enough to exceed the solubility limit for carbon in austenite. Coarse carbides are often found at corners and edges of parts made of alloys that are rich in strong carbide formers, such as chromium. Plate martensite and high levels of retained austenite can be found with coarse carbides in parts quenched directly from the carburizing temperature. Processing is sometimes designed to produce large primary carbides as a means of enhancing wear resistance (Ref 12). More often, large carbides are avoided, because they deplete the matrix in alloying elements such as Cr, thereby reducing hardenability.

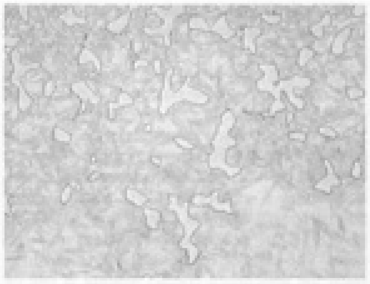


Fig. 4 Coarse primary carbides produced by carburizing SAE 4130 steel at 950 °C (1740 °F), and then quenching. Matrix microstructure is plate martensite and retained austenite. Picral etch. 600×

Carbide networks (Fig. 5) form in austenite grain boundaries when parts are carburized at an elevated temperature, and then slowly cooled. The solubility limit in austenite is exceeded as cooling occurs, and carbon is rejected to austenite grain boundaries. Because this structure tends to embrittle the case, it is usually avoided.

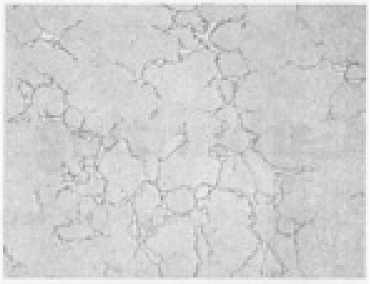


Fig. 5 Carbide networks in prior austenite grain boundaries. Produced by carburizing 4130 steel at 950 °C (1740 °F), furnace cooling to 800 °C (1470 °F), and then quenching. Picral etch. 600×

Fine primary carbides (from 0.1 to 0.5 μm, or 4 to 20 μin. diam) (Fig. 6) result when a part is carburized at a high temperature, such as 950 °C (1740 °F), cooled to form pearlite or bainite, and then reheated to a lower temperature, such as 830 °C (1525 °F), for a brief time and quenched. Because the carbon solubility at 950 °C (1740 °F) is on the order of

1.5 times the solubility at 830 °C (1525 °F), substantial quantities of fine carbides can be produced. The carbides will not coarsen significantly if the time at the lower austenitizing temperature is on the order of 30 min or less. Because the retained austenite content is a function of the carbon dissolved in austenite, the hardness will be near the maximum for the alloy. Finely dispersed primary carbides can incrementally increase the hardness. Their main contribution is in restricting austenite grain growth, thereby assuring fine martensite plates and finely dispersed retained austenite. Many gears and bearings are heat treated in this manner.

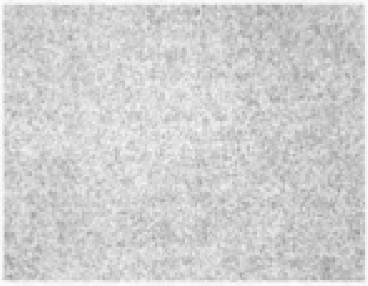


Fig. 6 Fine primary carbides in lath martensite produced by carburizing at 950 °C (1740 °F), air cooling to room temperature, then reheating to 820 °C (1510 °F) for 20 min and quenching. Picral etch. 600×

Residual Stress. Carburized parts usually contain compressive residual stresses in the high-carbon case. The stresses arise because of the sequence of transformations upon quenching (Ref 13). The higher the carbon content, the lower the temperature at which martensite begins to form. The lower the transformation temperature, the greater the volume increase upon transformation. Therefore, upon cooling, the core transforms first and the volume change is accommodated by plastic deformation of the austenitic case. When the case transforms, the martensitic core accommodates the volume increase in the case by elastic, rather than plastic, deformation. Consequently, compressive residual stresses in the case tend to be balanced by tensile residual stresses in the core.

Actual stress distributions can be more complex than the scenario above would suggest, because:

- Temperature gradients within the part cause transformation to begin below the case/core interface, and then progress inward and outward simultaneously
- Austenite that is retained upon quenching does not contribute to the case expansion

From a consideration of force balances, it follows that when the case is thin relative to the core, compressive residual stresses in the case will be high. If the case is thick relative to the core, compressive stresses in the case will be low and tensile residual stresses in the interior will be high. Compressive surface residual stresses are beneficial in applications involving torsional and bending loads, because the tensile stresses caused by these loads are a maximum at the surface. Compressive residual stresses are also beneficial in resisting rolling contact fatigue because they offset the tensile stresses responsible for initiating and propagating subsurface fatigue cracks (Ref 14, 15). However, in some parts, high compressive surface residual stresses can cause problems, such as case/core interface cracks at tooth tips in gears.

Residual stresses can be significantly reduced by tempering at temperatures up to 200 °C (390 °F), without greatly affecting either the hardness or retained austenite content (Ref 16). Changes in residual stress commonly occur during service (Ref 17).

Transformation of Retained Austenite. Under the action of elastic stresses or plastic deformation, retained austenite can transform to martensite (Ref 18, 19, 20). Transformation often occurs in service in parts that experience high contact loads. The surface hardness increases, and the resulting volume expansion increases the compressive residual stress at the surface. Carburized roller bearings typically contain from 30 to 40% retained austenite in the case. Lowering the retained austenite content usually has an adverse effect on rolling contact fatigue life (Ref 7, 21).

On the other hand, manufacturers of instrument bearings, who strive to maximize dimensional stability, make an effort to eliminate retained austenite (Ref 22). Similarly, in applications requiring the best resistance to indentation, the retained austenite content is kept low to maximize hardness. Ball bearings, for example, experience "point loading," whereas gears and roller bearings experience "line loading." Therefore, manufacturers of ball bearings, attaching more importance to indentation resistance, choose a through-hardened steel, SAE 52100, which is readily processed to yield from 5 to 10% retained austenite and hardness values greater than 62 HRC.

Surface Oxidation. Most carburizing is done in environments that contain some oxygen. An atmosphere capable of carburizing steel will also reduce iron oxides, but it often contains sufficient oxygen to oxidize alloying elements, such as Si, Mn, Cr, and V. During carburizing cycles of several hours duration, oxides of these elements will form at the surface and/or in grain boundaries intersecting the surface (Fig. 7).

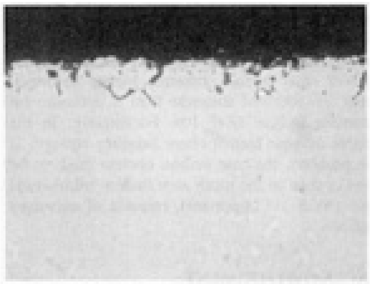


Fig. 7 Oxides formed at grain boundaries near the surface. Same sample as Fig. 5. Unetched. 600×

Depletion of these elements from the matrix can lower the hardenability enough to allow nonmartensitic products to form upon quenching, causing low hardness in a surface layer that is one to two grain diameters in thickness (Ref 23). Nonmartensitic products at the surfaces of gear teeth are known to reduce the tooth bending fatigue strength and would be expected to have an adverse effect on tribological properties, as well. If the matrix microstructure is unaffected, the effect of oxides on tribological properties is probably similar to that produced by oxide inclusions in the steel. Manufacturers of fine-pitch gears often do not attempt finishing operations to remove grain boundary oxides after heat treatment, arguing that the risk of surface damage caused by overheating during grinding outweighs any possible gain.

The thin layer of iron oxide that forms on carburized parts tempered in air at temperatures from 150 to 200 °C (300 to 390 °F) is usually benign. Thicker oxide layers, formed at higher temperatures, are sometimes used on steels as "break-in" coatings.

Carburizing Steels

Three factors are usually considered when selecting carburizing steels: cleanliness, hardenability, and fabricability, each of which is discussed below.

Cleanliness. For the best resistance to rolling contact fatigue (spalling), the content of aluminate, silicate, and globular oxide inclusions (Types B, C, and D, respectively, in the Jernkontoret system, ASTM E 45) must be as low as possible. Manganese sulfide inclusions (Type A) are generally not regarded as detrimental to rolling contact fatigue life (Ref 24).

The inclusion standards specified in ASTM A 534, "Carburizing Steels for Anti-Friction Bearings," have been steadily tightened since the late 1960s, reflecting improvements in steelmaking. Some individual steel suppliers claim to be able to furnish premium-quality carburizing steels with oxygen contents below 15 ppm, titanium contents below 30 ppm, and inclusion ratings considerably better than ASTM A 534. Calcium treatment of bearing-quality steels to modify aluminates, thereby improving machinability, is usually avoided, because of the possibility of forming large inclusions.

When sliding is combined with rolling contact, near-surface inclusions (including oxides formed in grain boundaries during heat treatment) promote pitting (Ref 25). The role of inclusions in most forms of sliding wear is not as well defined as their role in rolling contact fatigue, probably because the conditions that are possible at a sliding interface are more diverse and more difficult to characterize, than at a rolling contact interface. Some insight into possible effects of

inclusions on sliding wear comes from the machining literature (Ref 26). It is known that some inclusions in a steel workpiece can promote tool wear during machining, whereas others can reduce wear.

Hardenability. The alloy content of carburizing steels is usually selected on the basis of hardenability. If the application involves high contact loads (roller bearings, for example), the uncarburized core must be martensitic to prevent the subcase from yielding. An alloy that allows the part to attain full hardness (through-harden) in whatever quenchant is employed will be selected. The selection of an alloy with sufficient core hardenability almost always assures sufficient case hardenability. When contact loads are well within the capability of the case to support them, it is often not necessary nor desirable for the core microstructure to be martensitic. Shape distortion during quenching, for example, is usually reduced if the core transforms at a relatively high temperature to a nonmartensitic structure. For such parts, the alloying need only be sufficient to ensure case hardenability.

Jominy hardenability bands for alloy carburizing steels are described in the article "Hardenability Curves," in *Properties and Selection: Irons, Steels, and High-Performance Alloys,* Volume 1 of *ASM Handbook.* Narrower bands for restricted hardenability steels (Ref 27) have been devised in response to user demands for more precisely defined hardenability. Modern steelmaking practice provides much better control over hardenability, through composition control, than is reflected in the older standards.

The Jominy bands for carburizing steels provide a means of comparing the core hardenability of steels. To ensure adequate case hardenability, one needs a Jominy band for the same base composition, with the carbon content raised to about 0.4 wt%. (This presumes that one wishes all of the case with a carbon content of 0.4% or greater to be fully hardened.) For common alloys, such as 1524, 4023, 5120, 8620, and 8720, a "low-side estimate" of the case hardenability is given by the Jominy bands for 1541, 4042, 5140, 8640, and 8740, respectively. For other carburizing steels, case hardenability must be estimated from the composition or determined experimentally (Ref 28).

Fabricability. Manufacturing considerations also influence alloy selection. SAE 8620 steel readily transforms after hot rolling and annealing to a ferrite-pearlite microstructure (150-180 Brinell hardness) that has good machinability. If fabrication is by cold forming, rather than machining, then SAE 4118 might be selected, because it can be softened to a lower hardness with better ductility than 8620.

In manufacturing small automotive gears, steel rod is often annealed to facilitate blank-making by cold forming. Blanks are then normalized to a higher hardness to facilitate subsequent gear hobbing.

Unfortunately, with increasing hardenability comes increasing difficulty in softening. This is particularly true in alloys such as 4815 and 9310, in which the transformation to pearlite is very sluggish, requiring about one day at 600 °C (1110 °F) for completion. Because microstructures are usually less than optimum, machining of these alloys is more difficult.

Other Considerations. Several secondary hardening carburizing alloys have been developed for applications that require resistance to elevated temperatures, such as helicopter gearing and rock drill bits (Ref 29, 30, 31, 32). These alloys make use of the precipitation of copper and/or M_2C and MC carbides to provide resistance to softening for temperatures up to 550 °C (1020 °F). Because they contain substantial amounts of Mo and V, these alloys resemble low-carbon versions of tool steels. Some of the alloys are difficult to carburize because of high Si and Cr contents; preoxidation prior to carburizing is necessary to permit carbon penetration (Ref 33). Secondary hardening alloys could also be useful in ambient-temperature applications in which lubrication is marginal, because the heat generated by intermittent metal-to-metal contact would not readily soften the underlying metal (Ref 34). The article "Microstructures and Properties of Carburized Steels" in *Heat Treating,* Volume 4 of *ASM Handbook* (1991) expands upon the discussion in this section.

Process/Materials Selection for Wear Resistance

The independent variables available for controlling the microstructure/properties of carburized cases are those that define the carburizing alloy (composition, cleanliness), and those that define the carburizing process (time/temperature/carbon-potential carburizing history, time/temperature quenching history, time/temperature tempering history). These tools provide a considerable degree of control over these microstructural features:

Martensite

- (a) Carbon content of source austenite
- (b) Plate size (austenite grain size)
- (c) Strength
- (d) Secondary hardening

Primary carbides

- (a) Size
- (b) Volume fraction

Retained austenite

- (a) Volume fraction
- (b) Carbon content

Nonmetallic inclusions and these global features:

- Case depth
- Residual stress distribution

which determine the tribological properties of the case. The combination of properties that is best for each application must then be decided.

The necessary case depth and case hardness can be estimated from a Hertzian stress calculation, but other microstructural objectives can be specified only qualitatively. For many applications, the following "rules of thumb" apply:

- Sufficient case depth and case hardness must be provided to prevent indentation or case crushing under the anticipated contact loads. For gears and bearings loaded in "line contact," a minimum case hardness of 58 HRC frequently is specified. When high contact loads are accompanied by sliding, the near-surface hardness (to a depth of about 50 μm, or 2 mils) may have to be raised to prevent shearing of surface layers.
- The retained austenite content should be as high as possible, consistent with the requirements of the rule above. The retained austenite content should be controlled by adjusting the case carbon content, not by subzero quenching after carburizing or by tempering at temperatures above 200 °C (390 °F)
- The tempering temperature chosen should be as low as possible, but above the surface temperatures anticipated in finishing operations and in service
- The content of nonmetallic inclusions should be no higher than that needed for economical machining
- Coarse primary carbides can be helpful in resisting abrasive wear. Fine primary carbides can permit more retained austenite at the same hardness level. Experiments should be conducted to verify any benefits presumed to be associated with primary carbides

For certain types of gearing, the selection of materials and processes is dominated by concerns with resistance to tooth bending fatigue, rather than the tribological properties of the surfaces in contact. The best resistance to high-cycle bending fatigue appears to be associated with lower case carbon contents, higher hardness, and less retained austenite than is optimum for contact fatigue (Ref 10). Fortunately, in the roots of gear teeth (where bending strength is important), the case carbon content tends to be lower than in the tooth area (where tribological properties are important), because of curvature effects.

References

1. C.M. Kim, *Case Hardened Steels: Microstructural and Residual Stress Effects,* D.E. Diesburg, Ed., TMS-

AIME, 1984, p 59-87

2. C.A. Stickels, *ASM Handbook,* Vol 4, ASM International, 1991, p 312-324
3. G. Krauss, *Steels: Heat Treatment and Processing Principles,* ASM International, 1990
4. R.W.K. Honeycombe, *Steels: Microstructure and Properties,* American Society for Metals, 1981
5. K.-E. Thelning, *Steel and its Heat Treatment,* 2nd ed., Butterworths, London, 1984
6. C.A. Stickels, *Metall. Trans.,* Vol 5, 1974, p 865-874
7. J.A. Erickson, *Met. Prog.,* Vol 92, 1967, p 69-73
8. E.C. Bain and H.W. Paxton, *Alloying Elements in Steel,* American Society for Metals, 1966
9. W. Crafts and J.L. Lammont, *Hardenability and Steel Selection,* Pitman & Sons, London, 1949
10. C. Razim, *Alloys for the Eighties,* R.Q. Barr, Ed., Climax Molybdenum Co., 1980, p 9-23
11. H. Swahn, P.C. Becker, and O. Vingsbo, *Metall. Trans. A,* Vol 7A, 1976, p 1099-1110
12. R.F. Kern, *Heat Treat.,* Oct 1986, p 36-38
13. J.A. Burnett, *Residual Stress for Designers and Metallurgists,* American Society for Metals, 1981, p 51-69
14. G.T. Hahn, V. Bhargava, C.A. Rubin, and X.G. Leng, *Carburizing: Processing and Performance,* ASM International, 1989, p 101-113
15. D. Brooksbank and K.W. Andrews, *J. Iron Steel Inst.,* Vol 210, 1972, p 246-255
16. R.L. Brown, H.J. Rack, and M. Cohen, *Mater. Sci. Eng.,* Vol 21, 1975, p 25-34
17. H. Muro, N. Tsushima, and K. Nunome, *Wear,* Vol 25, 1973, p 345-356
18. F.T. Krotine, M.F. McGuire, L.J. Ebert, and A.R. Troiano, *ASM Trans. Quart.,* Vol 62, 1969, p 829-838
19. R.H. Richman and R.W. Landgraf, *Metall. Trans. A,* Vol 6A, 1975, p 955-964
20. C.A. Stickels, *Metall. Trans. A,* Vol 8A, 1977, p 63-70
21. H. Muro, Y. Sadaoka, S. Ito, and N. Tsushima, *Proceedings of the 12th Japanese Congress of Materials Research,* Mar 1969, p 74-77
22. T.J. Hughel, *ASM Trans. Quart.,* Vol 62, 1969, p 18-23
23. R. Chatterjee-Fischer, *Metall. Trans. A,* Vol 9A, 1978, p 1553-1560
24. P. Tardy, *Archiv Eisenhuttenw.,* Vol 43, 1972, p 583-587
25. T.M. Clarke, G.R. Miller, L.M. Keer, and H.S. Cheng, *ASLE Trans.,* Vol 28, 1985, p 111-116
26. F.W. Boulger, *Properties and Selection: Irons, Steels, and High-Performance Alloys,* Vol 1, 10th ed., *Metals Handbook,* ASM International, 1990, p 591-602
27. Standard J1868, *SAE Handbook,* Vol 1, Society of Automotive Engineers, 1990
28. Standard J406, *SAE Handbook,* Vol 1, Society of Automotive Engineers, 1990
29. C.F. Jatczak, *Met. Prog.,* Vol 113, 1978, p 70-78
30. W.E. Burd, *Met. Prog.,* Vol 127, 1985, p 33-35
31. R.A. Cutler and W.C. Leslie, *J. Test. Eval.,* Vol 11, 1983, p 3-15
32. J.D. Saulnier, *Carburizing: Processing and Performance,* ASM International, 1989, p 211-219
33. M.L. Schmidt, *J. Heat Treat.,* Vol 8, 1990, p 5-19
34. C.A. Stickels and C.M. Mack, *J. Heat Treat.,* Vol 4, 1986, p 223-236

Nitriding and Nitrocarburizing

F.T. Hoffmann and P. Mayr, Institute of Material Science (Germany)

Introduction

THE NITRIDING AND NITROCARBURIZING PROCESSES are, aside from carburizing, the most important thermochemical processes in heat treating industrial parts for the production of case-hardened surface layers. Layer of high hardness, which provide high resistance to corrosion and wear in addition to high fatigue strength, are produced by the diffusion of atomic nitrogen into the surfaces. Other advantages include very low distortion, because of the low temperatures involved and the absence of phase transformations, and high tempering resistance associated with the high hardness property at service temperatures below the nitriding temperature.

The original nitriding process used gaseous ammonia as the nitrogen source. Within the past 60 years, numerous nitriding and nitrocarburizing processes have been developed, some of them for specialized applications. Today, parts can be nitrided by using powders, salt baths, gaseous mixtures, amd plasma-assisted processes as nitrogen sources. Detailed information on the many forms of nitriding and nitrocarburizing are given in *Heat Treating,* Volume 4 of the *ASM Handbook* (1991).

Characteristics of Nitrided Surfaces

The structure of a nitrided layer comprises:

- A compound layer on the surface, consisting primarily of nitrides and carbonitrides of types Fe_4N and Fe_{2-3} (N,C). In addition, special nitrides are formed in steel that contain nitride-forming elements, such as aluminum, chromium, molybdenum, vanadium, and titanium. If specified by the user, the compound layer can be suppressed by a special nitriding practice
- A diffusion layer below the compound layer, composed of nitrogen in interstitial solution in the ferritic matrix in combination with nitride dispersions

The composition and microstructure of the nitride layer can be greatly affected by process selection and specific process parameters. The composition of the compound layer can be affected by the nitriding atmosphere. Higher temperatures or a longer nitriding time produce increased growth of the compound layer. The structure of the diffusion layer (that is, type, size, and distribution of the nitrides) can be affected by:

- The nitriding temperature
- The cooling rate from nitriding temperature to room temperature
- Low-temperature annealing after cooling to room temperature
- The type of steel being nitrided

High nitriding temperatures promote the formation of large, coarse distributed nitrides; low temperatures, promote the formation of finely dispersed nitrides. Slow cooling rates from nitriding temperature to room temperature, annealing treatments, or heating by stressing the parts favors dissolution, growth, and change in the nitride structure. The nitriding depth can be increased either by increasing the process temperature or by increasing the nitriding time. Increasing the content of nitride-forming elements in the material increases the hardness of the surface layer ad decreases the nitriding depth. Both the hardness and the depth parameters depend on the extent of nonferrous nitride formation.

In addition, the atmosphere composition and nitriding temperature (that is, the nitriding potential and the carbon potential of the atmosphere) affect the phase composition and the distribution of nitrogen and carbon in the compound layer. Investigations (Ref 1) have shown that if the carbon potential of a nitriding atmosphere is low, every nitriding process starts with decarburization of the matrix. Increasing carbon potential (when nitrocarburizing) and increasing nitriding potential (when nitriding) decreases decarburizing. This effect is present even when Armco iron is used. This leads to zones below the compound layer with a lower carbon content than in the matrix. The diffusion coefficient of carbon in the iron matrix is lowe than the diffusion coefficient of carbon in the nitride layer, but the solubility of carbon in the nitride layer is higher than the solubility of carbon in the iron matrix. Thus, after the formation of the compound layer, carbon diffusing to the surface is enriched in the compound layer at the interface with the iron matrix. Figure 1 is a plot of the compound layer composition as a function of nitriding time. The carbon content in this enriched zone depends on the type of forming nitride (that is, the nitriding potential of the atmosphere) and on the carbon content of the steel itself.

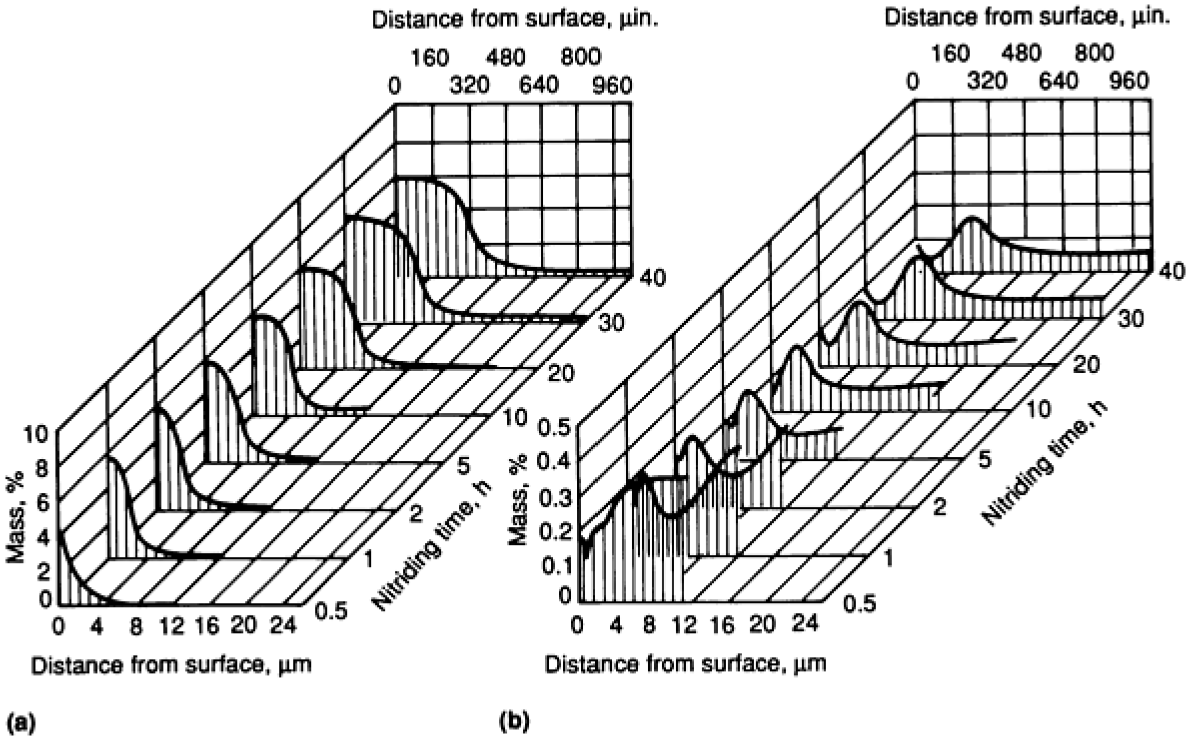


Fig. 1 Plot of compound layer concentration versus distance from the surface for Ck 45N nonresulfurized carbon steel as a function of gas nitriding duration. (a) Nitrogen distribution. (b) Carbon distribution. Nitrided at 540 °C (1005 °F). Nitriding number (K_N), 0.5

Wear Resistance of Nitrided and Nitrocarburized Materials

There are hundreds of articles in the literature that describe wear problems and their solution by different nitriding or nitrocarburizing processes. However, few of them describe the investigated nitride layers sufficiently. Most fail to specify the nitriding process, nitriding parameters, phase composition, compound layer thickness, and chemical composition and microstructure of the nitrided material. Therefore, the wear behavior of nitrided parts is somewhat difficult to ascertain from the literature. In this article, trends were deduced from recent, albeit contradictory, results (Ref 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33) in order to make generalized statements on wear behavior.

The wear resistance of a nitride layer changes with distance from the surface (Fig. 2). Wear resistance is reduced in the porous zone of the compound layer because of the lower fatigue strength, reduced density, and the notch effect of the pores. The wear resistance of the poreless zone of the compound layer is significantly higher than that of the diffusion zone and the core material. If the compound layer has a homogeneous structure, wear resistance is constant throughout. In the diffusion layer, wear resistance decreases to the value of the core material with increasing distance from the surface;

this is because of the reduction of the density of the nitride precipitates and the supersaturation of the matrix with nitrogen.

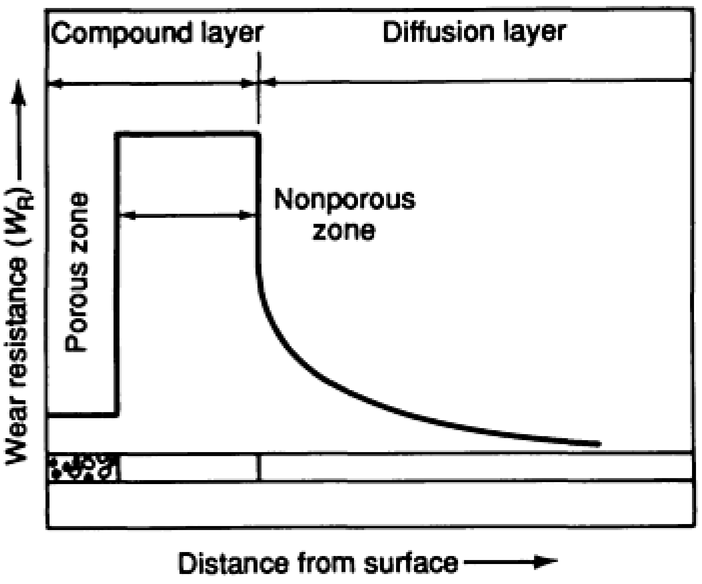


Fig. 2 Plot of adhesive and abrasive wear resistance versus distance from the surface for a nitride layer

The wear behavior of a nitride layer is often assumed to be an integral reaction of the nitriding layer to wear loading. This view does not explain wear mechanisms and their interaction during loading. Therefore, wear resistance and mechanisms must be discussed in conjunction with the structure of the nitriding layer. Table 1 gives the wear resistance of the compound layer and the diffusion layer in terms of the four major wear mechanisms discussed below.

Table 1 Effect of nitride layer on the wear resistance and the interaction parameter of metals as a function of wear mechanism

Wear mechanism	Interaction parameters		Relative wear resistance[b]	
	Compound layer[a]	**Diffusion layer[a]**	**Compound layer**	**Diffusion layer**
Adhesion	Structure	Formation of a protection layer (pl)	XXX	X
Tribo-oxidation	$v_{pl,form} > v_{pl,sol}$	$v_{pl,form} > v_{pl,sol}$	X	X
	$v_{pl,form} < v_{pl,sol}$	$v_{pl,form} < v_{pl,sol}$	[c]	[c]
Abrasion	Structure	Solid solution and precipitation hardening	XX	X

(a) $v_{pl,form}$, formation rate of friction-induced protective layer; $v_{pl,sol}$, solution rate of friction-induced protective layer.

(b) Increase relative to nonnitrided surface.

(c) Decrease in wear resistance

Adhesion. In adhesive wear, the wear resistance of the compound layer is very high compared with nonnitrided parts. This is because the inclination to microwelding decreases due to the change in the electron configuration in the outer region of the part. In the diffusion layer, wear resistance can also increase because of the formation of protective nitrogen-containing layers.

Tribo-oxidation. The wear resistance of the compound layer against tribo-oxidation depends on the formation of protective layers in the contact zone. If the formation rate, $v_{pl,form}$, of the friction-induced protective layer is greater than the concurrent solution rate, $v_{pl,sol}$, wear resistance increases. If $v_{pl,sol}$, > $v_{pl,form}$, wear resistance decreases. The wear resistance of the diffusion layer seems to be behave similarly.

Abrasion. The compound layer is very resistant to abrasive wear, because the structure of the compound layer allows only very low plastic deformation. Compared with nonnitrided steel, the wear resistance of the diffusion layer is higher--a result of the higher fatigue strength obtained by solid-solution strengthening and precipitation hardening.

Surface Fatigue. It is assumed that the wear resistance of nitrided parts to surface fatigue is higher than that of nonnitrided parts, as the changed lattice structure of the compound layer prevents plastic deformation. The diffusion zone has very high resistance to surface fatigue wear; here plastic deformations are very small because of solid-solution strengthening and precipitation hardening.

Influence of Variables on Wear Resistance of Nitrided Parts

Compound Layer. Variables that influence the wear resistance of the compound layer of nitrided parts are illustrated in Fig. 3, 4, 5, 6, 7, 8, and 9. A porous zone will raise the initial wear in the case of adhesive and abrasive wear (Fig. 3 and 4). If tribo-oxidation is the main wear mechanism, a large increase in wear resistance can occur (Fig. 5). This may be due to absorption of the lubrication medium and subsequent formation of a lubrication layer. In the case of surface fatigue (Fig. 6), compound layer thickness has no influence on wear resistance if the porous zone of the compound layer is small and the maximum of the true stress (σ_v) lies in the deeper regions of the compound layer or below it. If the thickness of the porous zone increases the distance from the surface of the maximum stress, wear resistance to surface fatigue will rapidly decrease.

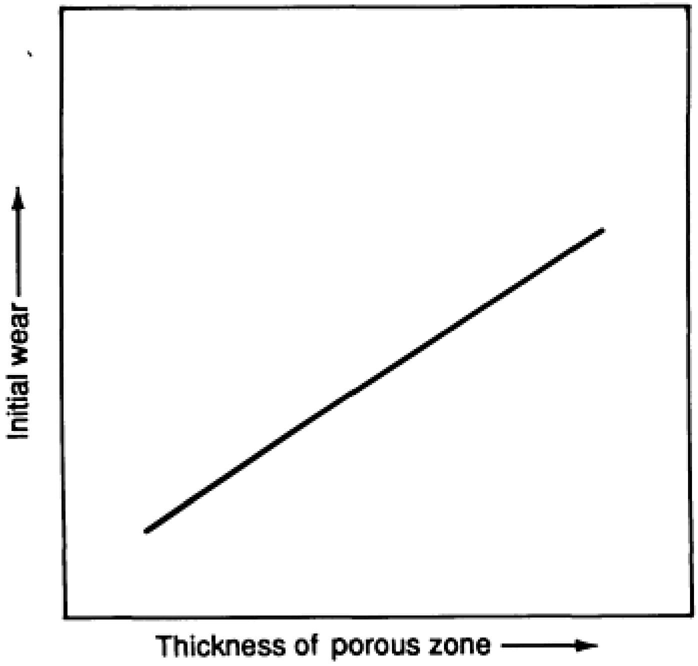


Fig. 3 Effect of porous zone thickness on adhesive wear resistance of the compound layer of a nitrided part

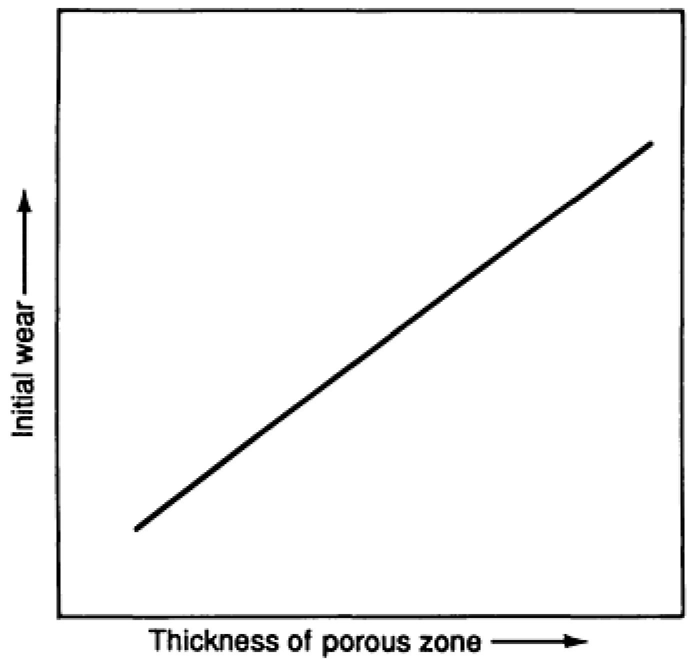


Fig. 4 Effect of porous zone thickness on abrasive wear resistance of the compound layer of a nitrided part

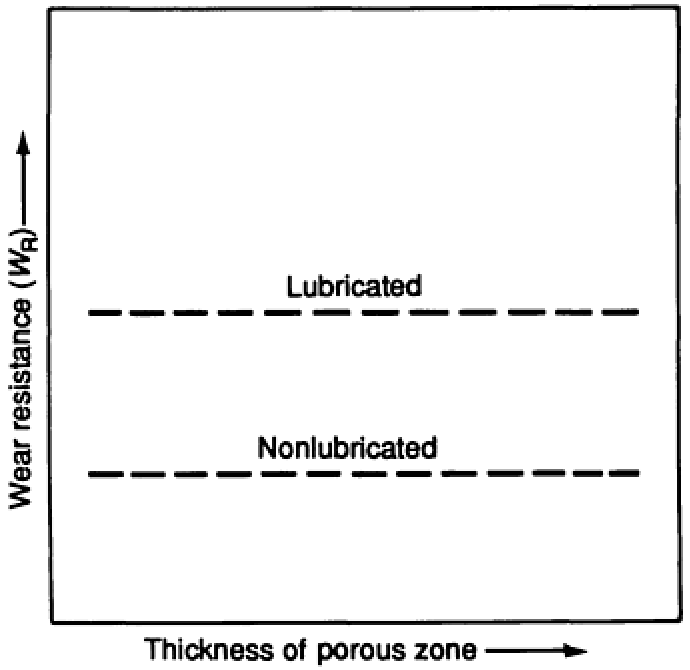


Fig. 5 Effect of lubrication on resistance of the compound layer of a nitrided part to wear caused by tribo-oxidation

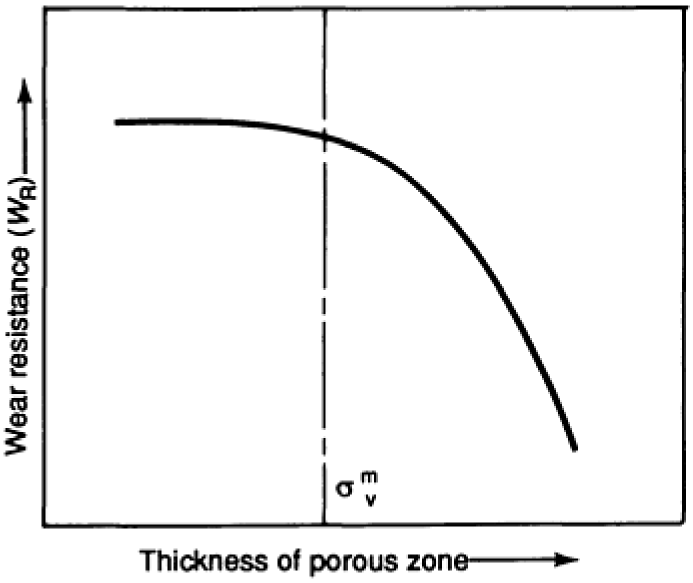


Fig. 6 Effect of porous zone thickness and maximum stress on resistance of the compound layer of a nitrided part to wear caused by surface fatigue

The structure and composition of the compound layer also influence its wear resistance. Investigations have shown that the adhesive wear resistance of the compound layer is strongly affected by the volume of ϵ-nitrides. In most cases, resistance increases with increasing ϵ-nitride content (Fig. 7). Similar behavior exists when nitrided parts are abrasively stressed (Fig. 8a). However, there is a difference between layers consisting only of nitrides and those containing carbonitrides. At constant volume of ϵ-nitride, carbonitride layers show significantly higher wear resistance than nitride layers (Fig. 8b). For surface fatigue, a higher content of ϵ-nitride in the compound layer will result increased wear resistance if the surface pressure is held constant (Fig. 9).

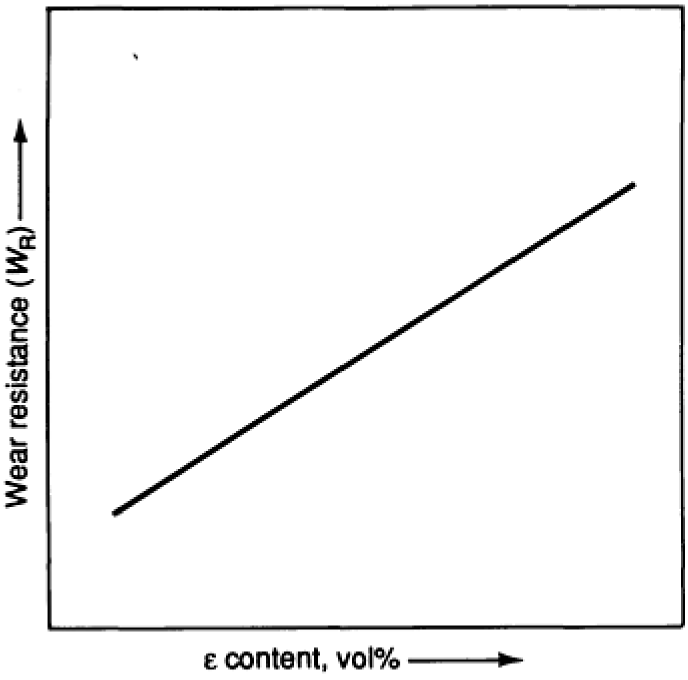


Fig. 7 Effect of ϵ-nitride content on adhesive wear resistance of the compound layer of a nitrided part

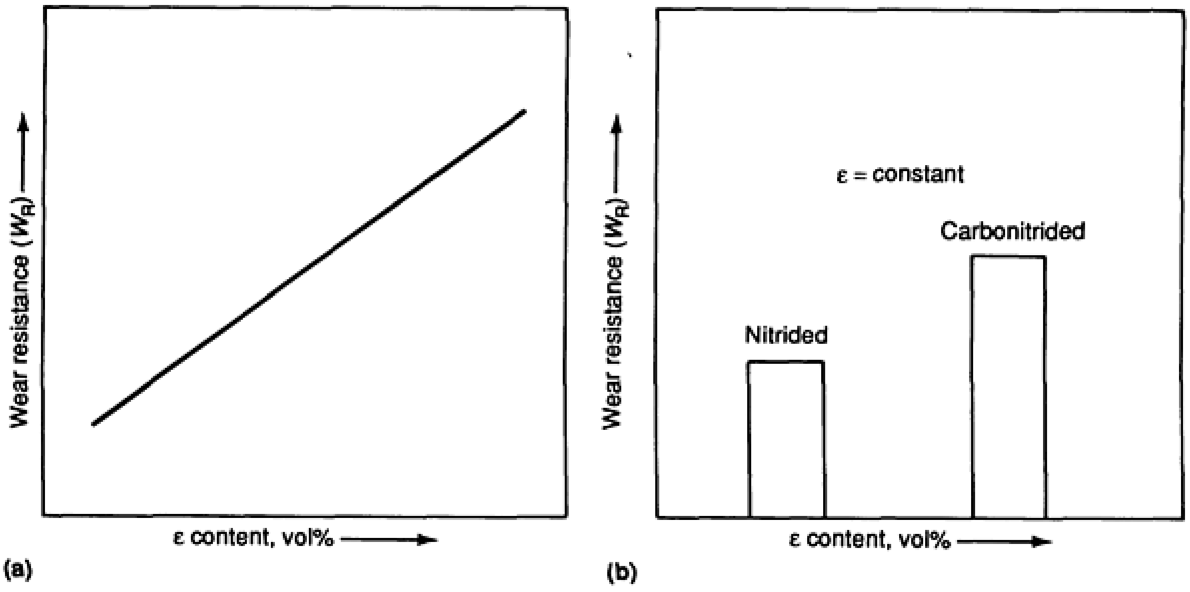


Fig. 8 Effect of ε-nitride content on abrasive wear resistance of the compound layer of a nitrided part. (a) Resistance increases with increasing εcontent. (b) At constant volume of ε, carbonitride layers show significantly higher resistance than nitride layers.

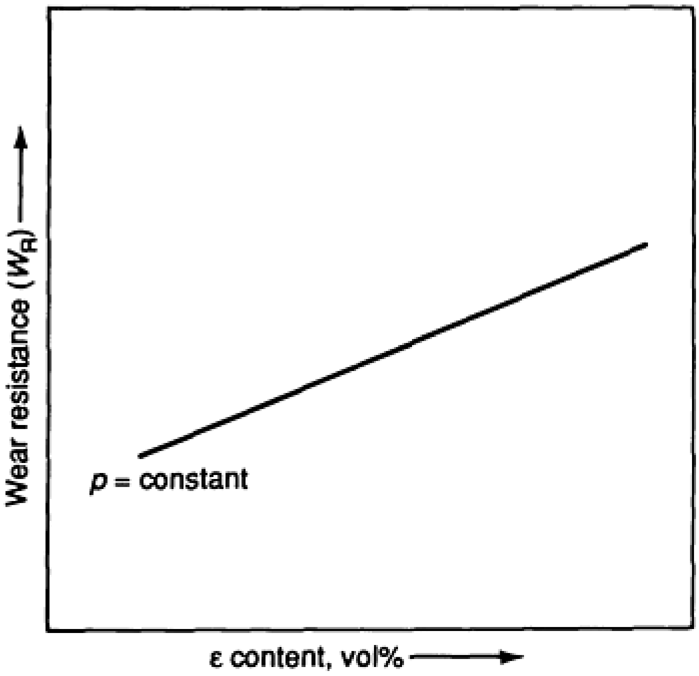


Fig. 9 Effect of ε-nitride content on resistance of the compound layer of a nitrided part to wear caused by surface fatigue at constant pressure

Diffusion Layer. Variables that affect the wear resistance of the diffusion layer of nitrided parts are illustrated in Fig. 10, 11, 12, 13, 14, 15, 16, and 17. Hardening of the alloy by precipitation methods such as solid-solution strengthening will raise the adhesive, abrasive, and surface fatigue wear resistance (Fig. 10 11 12). Greater nitriding depths also increase wear resistance (Fig. 13 14 15). Increases in the nitriding depth allow initial high wear resistance to abrasion to be upheld for a longer time, thus increasing component life (Fig. 14). Behavior is similar for surface fatigue stressing: increases in nitriding depth allow higher surface pressures and constant wear resistance (Fig. 15).

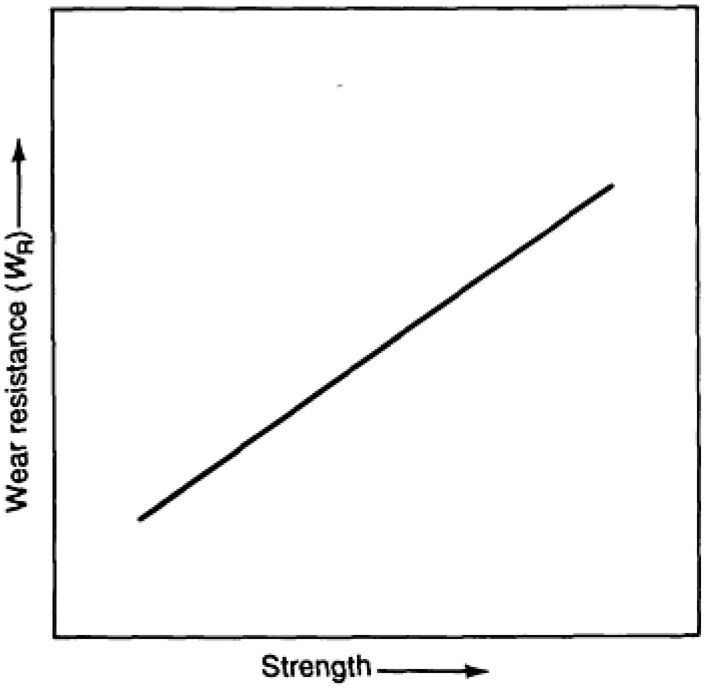


Fig. 10 Effect of material strength on adhesive wear resistance of the diffusion layer of a nitrided part

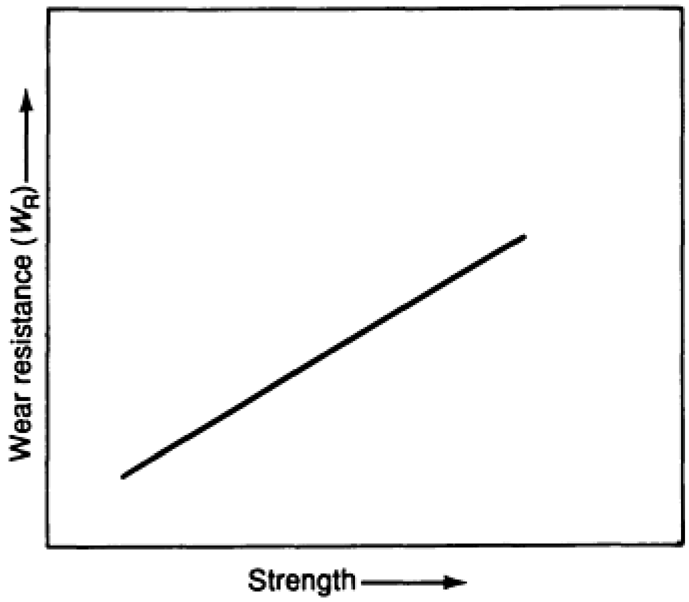


Fig. 11 Effect of material strength on abrasive wear resistance of the diffusion layer of a nitrided part

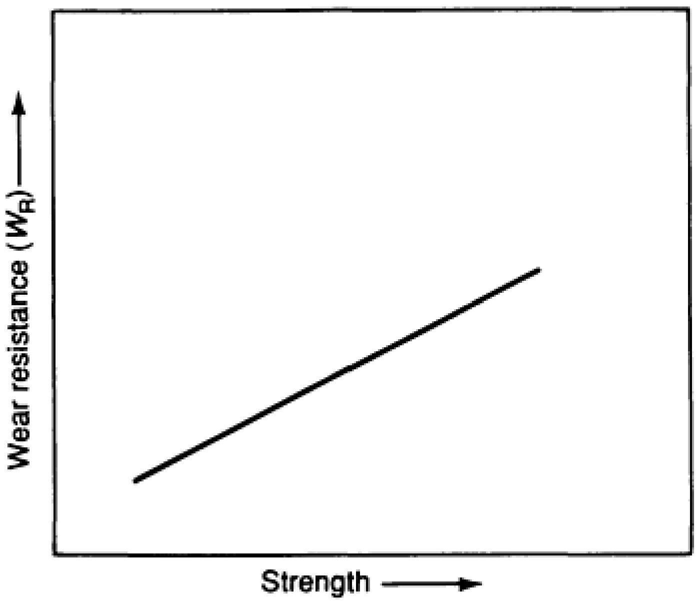


Fig. 12 Effect of material strength on resistance of the diffusion layer of a nitrided part to wear caused by surface fatigue

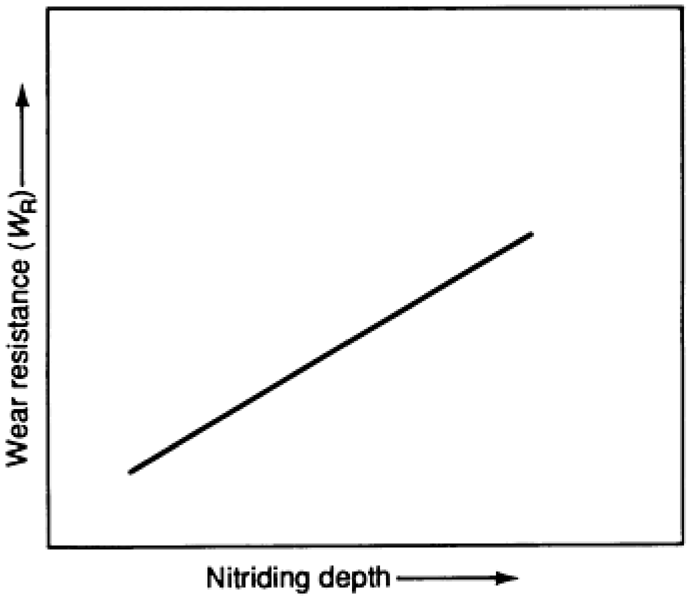


Fig. 13 Effect of nitriding depth on adhesive wear resistance of the diffusion layer of a nitrided part

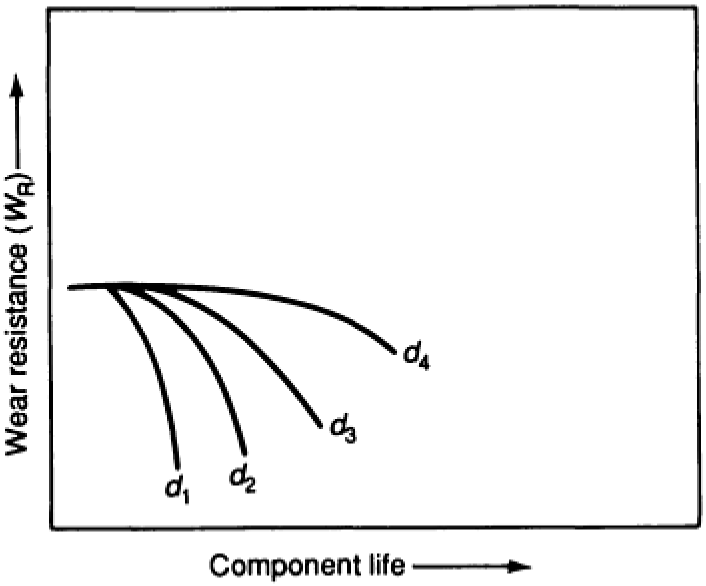


Fig. 14 Effect of nitriding depth on abrasive wear resistance of the diffusion layer of a nitrided part. Increases in nitriding depth increase component life. Nitriding depth, *d*: $d_1 < d_2 < d_3 < d_4$

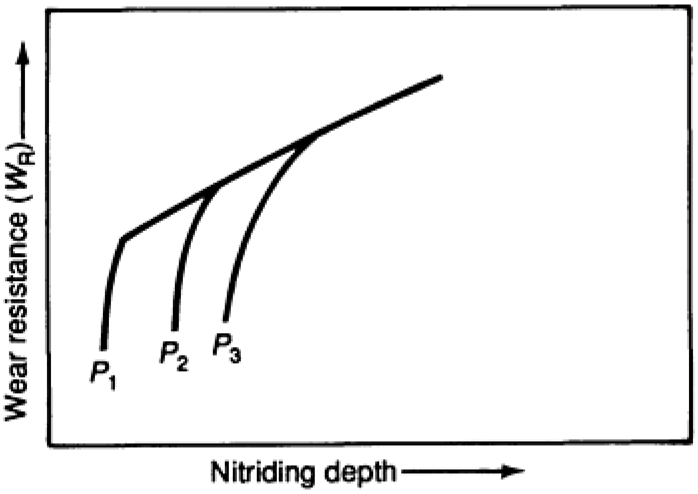


Fig. 15 Effect of nitriding depth on resistance of the diffusion layer of a nitrided part to wear caused by surface fatigue. Increases in nitriding depth allow the part to withstand higher surface pressures without sacrificing wear resistance. $p_1 < p_2 < p_3$

Nitride type and distribution in the diffusion layer can be changed by a postnitriding aging, strongly influencing wear resistance. For example, when nitrided parts are age hardened at constant temperature for different times, a characteristic hardness profile with a distinct maximum will result. Adhesive or abrasive wear stressing of such parts also results in a maximum wear resistance; however, the maximum shifts to higher aging times. Thus, both types of wear are sensitive to the type and distribution of the nitride precipitation in the diffusion layer: a matrix wit finely dispersed $Fe_{16}N_2$ nitrides has a lower wear resistance than one with a coarse distribution of Fe_4N nitrides (Fig. 16). The same is true for parts subjected to wear caused by surface fatigue (Fig. 17).

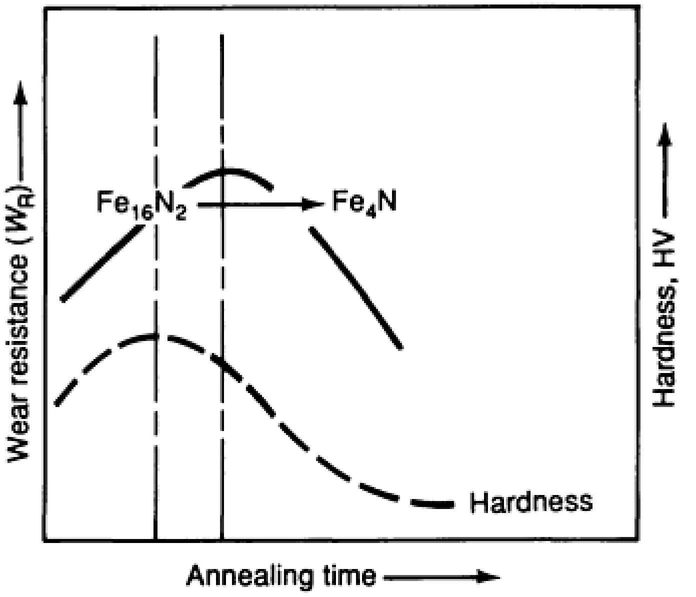


Fig. 16 Hardness profile showing effect of annealing time on abrasive and adhesive wear resistance of the diffusion layer of a nitrided part. See text for discussion of the effect of type and distribution of nitrides.

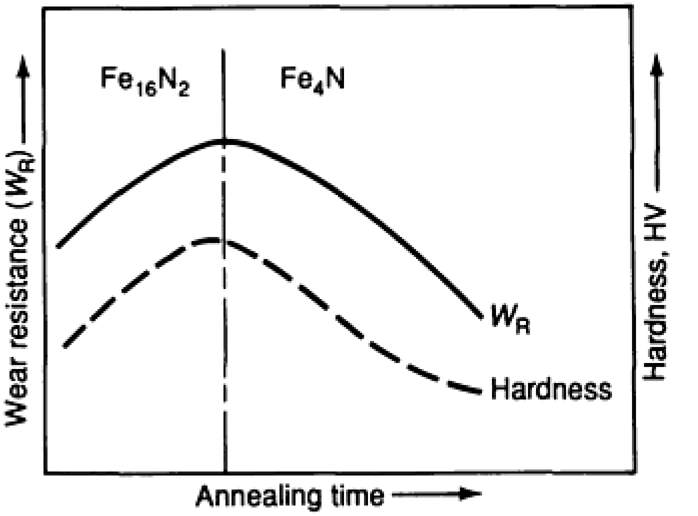


Fig. 17 Hardness profile showing effect of annealing time on resistance of the diffusion layer of a nitrided part to wear caused by surface fatigue. See text for discussion of the effect of type and distribution of nitrides.

Influence of the Nitriding/Nitrocarburizing Process on Wear Behavior

Investigations of the influence of the nitriding process on wear behavior can be divided into two groups: process oriented and materials science oriented. The objective of process-oriented investigations is often to show the clear advantages of a certain process. Materials-science-oriented investigations are usually more varied. Evaluation of published results of experimental work indicates that nitriding process variables do not exert much influence if nitriding layers are compared and are optimized to the actual wear stress.

Process variables do have an influence, however, if changes in geometry, which depend on the process, are considered. Typical process-dependent changes in geometry are shown in Table 2. Parts with elevations at their corners will have

shorter lifetimes because of the higher surface pressures at the corners than parts with constant surface pressures over the entire contact zone.

Table 2 Effect of nitriding process on surface topography and surface finish

Nitriding process	Material cross section showing surface topography	Nitriding parameters: Temperature °C (°F)	Nitriding parameters: Duration, h	Total roughness peak-to-valley (R_t), μm	Total roughness peak-to-valley (R_t), μin.	Arithmetic average surface roughness (R_a), μm	Arithmetic average surface roughness (R_a), μin.
Before nitriding		. . .	. . .	1	40	0.21	8.4
Plasma nitrided (bearing surface nitrided)		530 (985)	24	5	200	0.58	23
Plasma nitrided (totally nitrided)		550 (1020)	2	6	240	0.62	25
		550 (1020)	24	6	240	0.67	27
Salt-bath nitrided		570 (1060)	4	11	440	1.22	49
Gas nitrided		500 (930)	36	11	440	1.06	42
		500 (930)	84	13	520	1.21	48

Optimization of Wear by Process Technology

The influence of nitriding process parameters on wear resistance is not very high, but, with limitations, specific wear properties can be fitted to loading conditions by process optimizing. Table 3 summarizes possible enhancements of wear resistance for various wear mechanisms.

Table 3 General guidelines for improving the wear resistance of nitrided and carbonitrided steels as a function of wear mechanism and surface layer type

Layer	Wear mechanism(s)	Action required to improve wear resistance	Recommended change in process parameter
Compound layer	Abrasion, adhesion	Optimization of structure: • Transform γ' to ϵ microstructure by using higher nitriding potentials • Convert from nitriding to carbonitriding to add additional carbon donors • Minimize porous zone region by polishing surface	Modify nitriding atmosphere: • Salt bath: change content of base from CN^- to CNO^- • Gaseous: change nitriding number, flow rate, and carbon donors
		Formation of reactive layers	Add oxidation process after nitriding process is completed
	Tribo-oxidation	Modification of tribosystem	. . .
	Surface fatigue	Reduction in size of porous zone	Modify nitriding atmosphere
Diffusion layer	Abrasion, adhesion	Increase strength of material	Select alternate materials, add heat treatment prior to nitriding (use hardening and tempering operations instead

	Tribo-oxidation	Modification of tribosystem	. . .
	Surface fatigue	Optimization of hardness profile	Increase temperature and duration of nitriding process; select alternate materials; add heat treatment prior to nitriding; increase cooling rate after nitriding; add annealing treatment after nitriding
		Modification of composition (type, form, particle size, and distribution of nitrides) to	. . .

Adhesive or Abrasive Wear. In adhesive or abrasive wear, the properties of the compound layer must be optimized by:

- Changing γ'-nitride to ϵ-nitride by higher nitriding potentials (nitriding numbers)
- Changing nitrides to carbonitrides by adding carbon donors such as Endogas and CO_2 or by using a material with a higher carbon content
- Minimizing the porous zone by polishing

Porosity can also be influenced by the atmosphere (for example, lower nitriding potential or lower content of carbon donors), but this may worsen the properties of the compound layer. However, a small porous zone in the compound layer may be advantageous when the shape of the surface can be optimized by running-in (initial) wear.

In gas nitriding, the structure of the compound layer can be changed by a definite nitriding atmosphere--that is, a definite nitriding number [ratio $p(NH_3)/p(H_2)^{1.5}$] or by additional carbon-containing gases such as CO_2 or Endogas. In plasma nitriding, additional parameters such as pressure and glow discharge conditions are beneath the atmosphere. In bath nitriding, the concentration of the cyanate and cyanide content of the bath can be changed.

A reaction layer can be formed by an oxidation process after nitriding. Usually, these processes are performed in salt baths or in gaseous atmospheres.

The adhesive and abrasive wear resistance of the diffusion layer can be enhanced by strengthening the diffusion layer. The most effective means of this are by proper material selection (for example, high content of nitride-forming elements) and by heat treatment before nitriding (for example, hardening and tempering instead of normalizing).

Tribo-oxidation. If tribo-oxidation is the primary wear mechanism, only limited improvements can be made by changing the nitriding conditions. In this case, it is necessary, if possible, to change the conditions of the tribosystem.

Surface Fatigue. If controlling surface fatigue is the objective, measures to optimize the diffusion layer can be effective--in particular, fitting the nitriding hardness profile to the stressing profile and influencing the nitride precipitates (type, form, size, and distribution). The choice of material and heat treatment before nitriding depend on the required core strength. Nitriding of alloyed steels often results in thick deposits of carbides and/or nitrides in the grain boundaries, which may promote crack initiation. If the parts are to be under high load, such deposits should be avoided by changing the heat treatment process.

The nitriding depth can be increased by increasing process temperature and/or treating time. If merely prolonging the nitriding time does not result in sufficient nitriding depths, a two-step process can be performed. In the first step, nitriding is done at low temperatures to generate a finely dispersed nitride distribution. The second nitriding step is done at higher temperatures to raise the nitriding depth. This procedure avoids coarsening of the nitride distribution and thus a decrease in hardness. The cooling rate from the nitriding temperature and postnitriding treatments can be used to influence the type, size, and distribution of the nitrides.

Corrosive Wear. Nitriding improves resistance to corrosive wear. Resistance can be further improved by oxidation of the compound layer after nitriding. To achieve a very smooth layer capable of sustaining high loads, it is recommended that the oxidized layer be polished and then oxidized again.

References

1. H. Klümper-Westkamp, F. Hoffmann, and P. Mayr, *Härt.-Tech. Mitt.,* Vol 44, 1989, p 346-355
2. K.H. Zum Gahr, Reibungs- und Verschleissmodelle, *Reibung und Verschleiss, Mechanismen-Prüftechnik-Werkstoffeigen-schaften,* K.H. Zum Gahr, Ed., DGM, 1983, p 53-78
3. "Verschleiss, Begriffe, Systemanalyse von Verschleissvorgängen, Gliederung des Verschleissgebietes," DIN 50 320, 1979
4. K.H. Habig, Systemtechnik tribologischer Vorgänge, *Reibung und Verschleiss, Mechanismen-Prüftechnik-Werkstoffeigen-schaften,* K.H. Zum Gahr, Ed., DGM, 1983, p 13-28
5. A. Begelinger and A.W.S. de Gee, TNO Delft Report, 1967
6. A. Begelinger and A.W.S. de Gee, *Lubr. Eng.,* Vol 16, 1970, p 56-63
7. H. Mittmann and H. Czichos, *Materialprüfung,* Vol 17, 1975, p 366-372
8. J. Föhl, *Reibung und Verschleiss, Mechanismen-Prüftechnik-Werkstoffeigen-schaften,* K.H. Zum Gahr, Ed., DGM, 1983, p 29-51
9. K.-H. Habig, *Verschleiss und Härte von Werkstoffen,* Carl Hanser Verlag, 1980
10. K.-H. Habig, Möglichkeiten der Model-Iverschleissprüfung, *Werkstoffe und ihre Veredlung,* Vol 2, 1980, p 229-232
11. W. Schröter, W. Uhlig, and G. Alisch, *Schmierungstechnik,* Vol 11, 1980, p 9-15
12. A.E. Miller, *J. Met.,* 1983, p 56-62
13. K.H. Zum Gahr, *Reibung und Verschleiss, Mechanismen-Prüftechnik-Werkstoffeigen-schaften,* K.H. Zum Gahr, Ed., DGM, 1983, p 135-156
14. D. Liedtke, *Z. Wirstsch. Fertigung,* Vol 65, 1970, p 234-237
15. J.C. Gregory, *Tribology Int.,* Vol 3, 1970, p 73-83
16. K.-H. Habig, R. Chatterjee-Fischer, and F. Hoffmann, *Härt.-Tech. Mitt.,* Vol 33, 1978, p 28-35
17. K.-H. Habig, W. Evers, and R. Chatterjee-Fischer, *Härt.-Tech. Mitt.,* Vol 33, 1978, p 272-280
18. C. Gleave and M. Farrow, *Heat Treatment '81,* The Metals Society, London, 1981, p 123-129
19. K.-H. Habig, "Verschleissuntersuchungen an gas-, bad- und ionitriertem Stahl 42 CrMo 4," BAM-Berichte No. 38, June 1976
20. A. Dubus and J.-P. Peyre, *Trait. Therm.,* 1984, p 27-31
21. F. Hoffmann, Lecture 38, Härtereikolloquium, 6-8 Octo 1982
22. D. Römpler, *Härt.-Tech. Mitt.,* Vol 34, 1979, p 219-226
23. P. Hammer and G. Polzer, *Fertigungstech. Betr.,* Vol 15, 1965, p 498-500
24. J. Zysk, *Härt.-Tech. Mitt.,* Vol 31, 1976, p 137-144
25. Fr.W. Eysell, *Durferr. Hausmitt.,* Vol 28, 1958, p 7-18
26. H. Tauscher and E. Stecher, *IfL-Mitt.,* Vol 4, 1965, p 220-224
27. H. Tauscher and H. Fleicher, *Maschinenbautechnik,* Vol 19, 1979, p 201-205
28. T.M. Norn and L. Kindbom, *Stahl Eisen,* Vol 78, 1958, p 1881-1891
29. D. Edenhofer, *Fachber. Oberflächentech.,* Vol 12, 1974, p 97-102
30. C.S. Nanjunda Ram and A. Ramamohana, *Materialprüf.,* Vol 1981, p 125-128
31. K. Keller, *Ind. Anzeig.,* Vol 8.09, 1967, p 1580-1581
32. R. Woska, *Härt.-Tech. Mitt.,* Vol 38, 1983, p 10-17
33. H. Döpke, *Z. Wirtsch. Fertigung,* Vol 71, 1981, p 273-274

Glossary of Terms

Compiled by Peter J. Blau, Metals and Ceramics Division, Oak Ridge National Laboratory

WORDS AND PHRASES are the essential tools of technical communication. Without clearly defined terminology, technological progress is impossible. The written and spoken word must be precise and mutually understood by all communicating individuals in the same context. Unfortunately, the terminology used in friction, lubrication, and wear science and technology ("tribology") has evolved from sources in many different scientific and engineering fields. Some terms (especially hyphenated terms) are coined during wear failure analysis and do not necessarily conform to existing terminology. Consequently, the terminology varies greatly in origin and breadth of accepted usage. Terms also vary in specificity; that is, many terms with slightly different connotations are sometimes erroneously substituted for one another. For example, in certain instances the surface of a plain bearing might be described alternately as being polished, burnished, or scuffed. Strictly speaking, these three terms are not equivalent. The use of tribology terminology also varies between different English-speaking countries. A term such as *density* can be rather clearly and unambiguously defined, but scuffing cannot because too many different interpretations of the term are in common use. To avoid confusion and misunderstanding, terms which are known to have several possible interpretations should not be used.

Any compilation of a glossary of terms in tribology becomes a difficult task because, to be truly successful, it should remove ambiguity and form a basis for future usage. However, the glossary must also allow users to find definitions for terms that are deeply entrenched in the literature of the field even though they might be vague or inconsistently used. Avoiding the pit-falls of providing too narrow a series of definitions and descriptions of terms necessitated using a variety of references in assembling this glossary. The definitions obtained from these references are supplemented by additional ones written specifically for this Volume in order to provide more complete coverage. Sometimes the definitions presented in the cited references were inconsistent with the style or intent of this glossary and were edited. When such editing was necessary, the reference number is followed by the letter "m." More than one entry was provided for certain terms in order to show differences in common usage. In addition to the definitions given here, the reader is referred to articles dealing with specific fields where additional terminology is provided (for example, the parts of a bearing or the types of lubrication).

The basic form of an entry in the glossary contains the word or phrase in boldface type, one or more definitions and/or notes, a reference number (no reference number is given if the definition was written for this glossary), and additional information about synonyms and/or related terms as follows:

- **term in boldface type.**
  - (1) First definition. *Note:* a note contains supplementary clarifying material not required in the formal definition (Ref *X*). (2) Another definition, possibly from another source (Ref *Y*). Synonym or see also *another term*

Not all terms in the cited references were included here for various reasons. For example, a given term might occasionally be used in tribology, but is not principally associated with friction, lubrication, and wear (for example, *belleville spring*), or the definition from one source was nearly identical to that from another and need to be repeated.

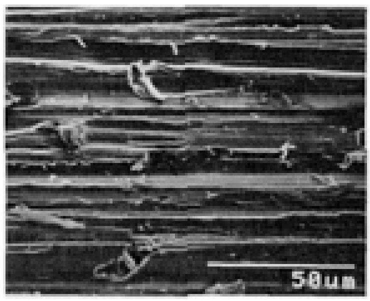
While the editors have attempted to coordinate the use of terminology throughout this Volume, we can offer no guarantee that the terminology used in the other Sections of this Handbook is entirely consistent with the definitions presented in this glossary. As with friction, lubrication, and wear technology as a whole, this Volume was compiled by individuals with backgrounds in many different fields.

To further clarify certain terms in this glossary, illustrations or photographs are included. Photographs, like written language, leave some room for precision, and examples were chosen to represent the general case of the phenomenon, not necessarily the only way in which a certain tribological phenomenon can appear.

- **A**
- **a-spot**
  - One of many small contact areas through which electrical current can pass when two rough, conductive, solid surfaces are touching. *Note:* This term derives from the analysis of R. Holm in

modeling electrical contact behavior. A-spots may represent an area smaller than the total asperity contact area referred to in other surface contact models.

- **Abbott-Firestone curve**
  - Introduced in 1933 by E.J. Abbott and F.A. Firestone and sometimes called a "bearing area curve," the Abbott-Firestone curve is a plot of depth below a reference level parallel to a solid surface (ordinate) versus the percent bearing area intercepted by a horizontal line at that depth (abscissa). *Note:* The Abbott-Firestone curve is sometimes used to compare the load-supporting behavior of bearings with different surface roughnesses. See also bearing area .
- **abrasion**
  - A process in which hard particles or protuberances are forced against and moving along a solid surface. *Note:* Sometimes this term is used to refer to abrasive wear . See also abrasive erosion .
- **abrasive erosion**
  - Erosive wear caused by the relative motion of solid particles which are entrained in a fluid, moving nearly parallel to a solid surface (Ref 1m). See also erosion
- **abrasive wear**
  - Wear by displacement of material caused by hard particles or hard protuberances (Ref 1). Wear due to hard particles or hard protuberances forced against and moving along a solid surface (Ref 2).



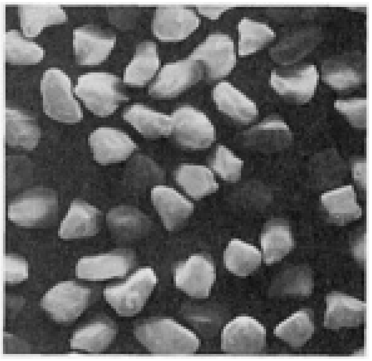
Abrasive wear of the surface of 1020 steel abraded by 220 grit SiC paper showing characteristic grooves and attached, tiny cutting chips. Courtesy of L.K. Ives, NIST

- **abrasive wear factor**
  - An empirical factor (f_v) that expresses the influence of various service environments on the increase in the radial clearance of a bearing (*V*) and, hence, the extent of life reduction as:

$$f_v = (V/e_0)$$

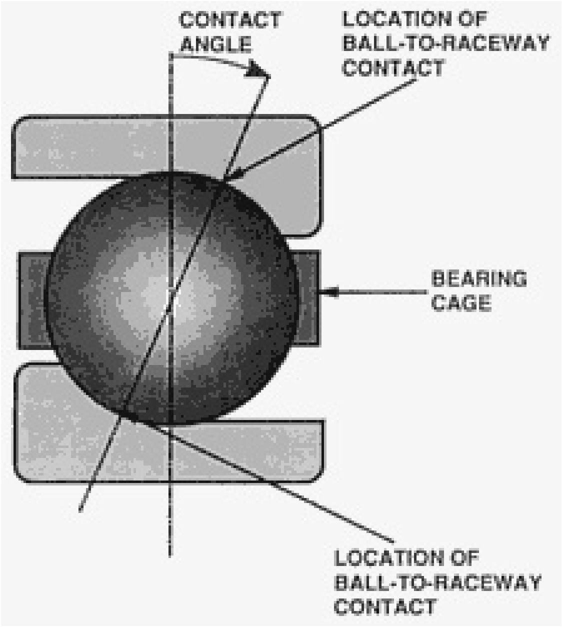
  - where e_0 is a geometrical factor (Ref 3, p 699).
- **abrasivity**
  - The extent to which a surface, particle, or collection of particles will tend to cause abrasive wear when forced against a solid surface under relative motion and under prescribed conditions. *Note:* Specifically, an ASTM standard for the abrasivity of slurries--ASTM G 75--has been developed.
- **absolute impact velocity**
  - See impact velocity .
- **asolute viscosity**
  - See viscosity .
- **acceleration period**
  - In cavitation and liquid impingement erosion, the stage following the incubation period, during which the erosion rate increases from near zero to a maximum value (Ref 2).
- **accessory seal**
  - On various types of engines, a seal that is employed for sealing an accessory shaft in the gear box, such as a shaft for operating an oil pump, a fuel pump, a generator, a starter, or a de-oiler (Ref 4m).
- **accumulation period**

- See preferred term acceleration period .
- **actual contact area**
  - The total area of contact formed by summing the localized asperity contact areas within the apparent area of contact . Also known as real area of contact .
- **actual slip**
  - See macroslip .
- **additive**
  - In lubrication, a material added to a lubricant for the purpose of imparting new properties or of enhancing existing properties. *Note:* Main classes of additives include anti-corrosive , antifoam , antioxidant , antiwear , detergent , dispersant , extreme-pressure , and VI improver additives (Ref 1m).
- **adherence**
  - In tribology, the physical attachment of material to a surface (either by adhesion or by other means of attachment) that results from the contact of two solid surfaces undergoing relative motion. *Note:* Adhesive bonding is not a requirement for adherence because mechanisms such as mechanical interlocking of asperities can also provide a means for adherence. See also adhesion (adhesive force) and adhesion, mechanical .
- **adhesion (adhesive force)**
  - In frictional contacts, the attractive force between adjacent surfaces. *Notes:* In physical chemistry, adhesion denotes the attraction between a solid surface and a second (liquid or solid) phase. This definition is based on the assumption of a reversible equilibrium. In mechanical technology, adhesion is generally irreversible. In railway engineering, adhesion often means friction (Ref 1).
- **adhesion, mechanical**
  - Adhesion between surfaces produced solely by the interlocking of protuberances on those surfaces. See also adherence .
- **adhesion coefficient**
  - See coefficient of adhesion .
- **adhesive wear**
  - (1) Wear by transference of material from one surface to another during relative motion due to a process of solid-phase welding. *Note:* Particles that are removed from one surface are either permanently or temporarily attached to the other surface (Ref 1). (2) Wear due to localized bonding between contacting solid surfaces leading to material transfer between the two surfaces or loss from either surface (Ref 2).
- **aerodynamic lubrication**
  - See gas lubrication .
- **aerostatic lubrication**
  - See pressurized gas lubrication .
- **AFBMA**
  - Abbreviation for Anti-friction Bearing Manufacturers' Association, an organization that produces specifications, such as the grading system for the quality of bearing geometry and finish, for the bearing industry.
- **AFS 50-70 test sand**
  - A rounded quartz sand specified for use as an abrasive in the dry sand-rubber wheel abrasive wear test (ASTM G 65). *Note:* Using the U.S. Sieve Series, none of this sand will be retained on Sieve No. 40, 5% maximum will be retained on Sieve No. 50, 95% minimum will be retained on Sieve No. 70, and none will pass Sieve No. 100. This places all the particle diameters between 425 and 150 μm.



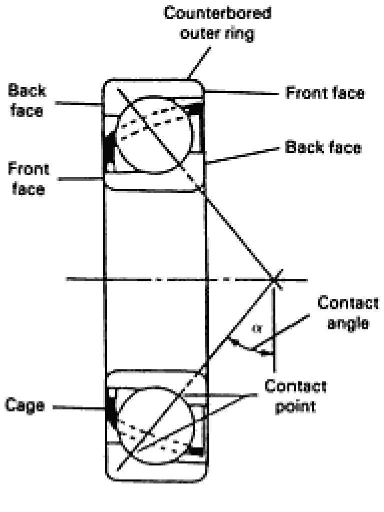
AFS 50-70 test sand particles used in the ASTM dry sand-rubber wheel abrasion test. Courtesy of L.K. Ives, NIST

- **age hardening (of grease)**
  - The increasing consistency of a lubricating grease with time of storage (Ref 1).
- **air bearing**
  - A bearing using air as a lubricant (Ref 1). See also gas lubrication and pressurized gas lubrication .
- **air hammer**
  - In lubrication, a type of instability, basically a resonance, that occurs in externally pressurized gas bearings (Ref 1).
- **aligning bearing**
  - A bearing with an external spherical seat surface that provides a compensation for shaft or housing deflection or misalignment (Ref 1). Compare with self-aligning bearing .
- **Alpha Model 1**
  - A type of wear-testing machine consisting of a conforming or flat-faced block pressed vertically downward by a deadweight loading arrangement against the circumference of a hardened steel ring that is rotating on a shaft. *Note:* This term is based on manufacturer's designation, but has become common terminology for referring to the block-on-ring geometry. Further information can be found in ASTM Standard Test Method G 77.
- **Amontons' laws**
  - Two laws propounded in 1699 that state that (1) friction force is proportional to normal force, and (2) friction force is independent of the size of the contact area (Ref 1).
- **Amsler wear machine**
  - A wear and traction-testing machine consisting of two disk-shaped specimens oriented such that their axes are parallel and whose circumferential, cylindrical surfaces are caused to roll or roll and slide against one another. *Note:* The rotation rates of each disk may be varied so as to produce varying degrees of sliding and rolling motion.
- **angle of attack**
  - The angle between the direction of motion of an impinging liquid or solid particle and the tangent to the surface at the point of impact (Ref 2).
- **angle of contact**
  - In a ball race, the angle between a diametral plane perpendicular to a ball-bearing axis and a line drawn between points of tangency of the balls to the inner and outer rings (Ref 1).



Angle of contact in an angular-contact bearing

- **angle of incidence**
  - The angle between the direction of motion of an impinging liquid or solid particle and the normal to the surface at the point of impact (Ref 2).
- **angular-contact bearing**
  - A ball bearing of the grooved type, designed in such a way that when under no load, a line through the outer and inner raceway contacts with the balls to form an angle with a plane perpendicular to the bearing axis (Ref 1).



Angular-contact bearing.

- **annular bearing**
  - (1) Usually a rolling bearing of short cylindrical form supporting a shaft carrying a radial load (Ref 1). (2) A flat disk-shaped bearing (Ref 1).
- **aniline point**
  - As applied to a petroleum product, the lowest temperature at which the product is completely miscible with an equal volume of freshly distilled aniline. *Note:* The aniline point is a guide to the oil composition (Ref 1).

- **anticorrosive additive**
  - A lubricant additive used to reduce corrosion (Ref 1m).
- **antiextrusion ring**
  - A ring that is installed on the low-pressure side of a seal or packing, in order to prevent extrusion of the sealing material (Ref 4).
- **antifoam additive**
  - An additive used to reduce or prevent foaming (Ref 1m). Also known as foam inhibitor .
- **antifriction bearing**
  - A bearing containing a solid lubricant (Ref 1m). *Note:* Historically, the term antifriction arose from the relatively slow starting torque of roller bearings compared with sliding bearings. See also roller bearing , rolling-element bearings, and self-lubricating bearing .
- **antifriction material**
  - A material that exhibits low-friction or self-lubricating properties.
- **antioxidant**
  - Any additive for the purpose of reducing the rate of oxidation and subsequent deterioration of oils or greases (Ref 5).
- **antiscuffing lubricant**
  - A lubricant that is formulated to avoid scuffing. *Note:* Antiscuffing lubricant has been suggested as a more accurate alternative to the term extreme-pressure lubricant .
- **antiseizure property**
  - The ability of a bearing material to resist seizure during momentary lubrication failure (Ref 1).
- **antiwear additive**
  - A lubricant additive used to reduce wear (Ref 1m).
- **antiwear number**
  - (AWN). A variant of the wear coefficient (*k*) suggested by Archard and defined as follows:

$$\text{AWN} = -\log_{10} k$$

  - See also Archard wear law .
- **antiweld characteristic**
  - See antiseizure property .
- **API gravity (API degree)**
  - A measure of density used in the U.S. petroleum industry where (Ref 1):

$$\text{Degrees API} = [141.5/\text{specific gravity at } 60\ ^\circ\text{F}) - 131.5$$

- **Apiezon oil**
  - An oil of low vapor pressure used in vacuum technology (Ref 1m).
- **apparent area of contact**
  - In tribology, the area of contact between two solid surfaces defined by the boundaries of their macroscopic interface (Ref 2). Contrast with real area of contact . See also Hertzian contact area and nominal contact area .
- **Archard wear law**
  - Introduced in 1953 by J.F. Archard, the Archard wear law expresses a proportionality among the sliding wear volume (*V*), the normal load (*W*), the total sliding distance (*S*), and the hardness (*H*) of the wearing surface:

$$V = \frac{kWS}{3H}$$

  - where *k*, the constant of proportionality, is often referred to as the "Archard wear constant," the "Archard wear coefficient," or simply the "wear coefficient" or the "wear constant." The use of the word "constant" in this context is not recommended. In some instances, the denominator is replaced with the flow pressure (*P*). *Note:* The simplicity of this linear law does not inherently

address the effects of sliding velocity, environment, or nonlinear sliding behavior associated with running-in and other wear transitions. See also wear coefficient , wear constant , and wear factor .

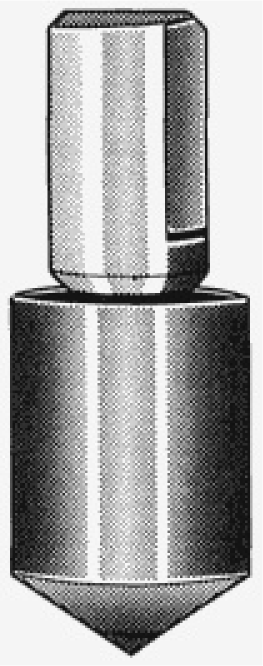
- **area of contact**
  - A general term that, without other modifying terminology, is insufficiently specific to be defined precisely. Its use should therefore be avoided. Related terms are actual area of contact, apparent area of contact , Hertzian contact area , nominal area of contact , and real area of contact .
- **asperities**
  - (1) The small-scale irregularities on a surface (Ref 1). (2) Minute imperfections on a seal face or surface of a mating ring that are the result of normal surface finishing processes (Ref 4). *Note:* Asperities, their shapes, sizes, and mechanical properties are the basis for developing many theoretical models for friction, lubrication, and wear behavior.
- **asperity**
  - In tribology, a protuberance in the small-scale topographical irregularities of a solid surface (Ref 2).
- **ASTM viscosity-temperature equation**
  - The equation relating kinematic viscosity (v) with temperature according to:

$$\log \log (v + 0.6) = m \log T + C$$

  - where v is the kinematic viscosity in centistokes, T is the absolute temperature, m is the ASTM slope (slope of the temperature-viscosity curve), and C is a constant that depends on the lubricant (Ref 1m).
- **atomic wear**
  - Wear between two contacting surfaces in relative motion attributed to migration of individual atoms from one surface to the other (Ref 1).
- **attenuation period**
  - See preferred term deceleration period .
- **attitude (attitude angle)**
  - In a bearing, the angular position of the line joining the center of the journal to that of the bearing bore, relative to the direction of loading (Ref 1m).
- **attrition**
  - Removal of small fragments of surface material during sliding contact (Ref 1).
- **average erosion rate**
  - The cumulative erosion divided by the corresponding cumulative exposure duration, that is, the slope of a line from the origin to a specified point on the cumulative erosion-time curve (Ref 2).
- **axial load bearing**
  - See thrust bearing .
- **axial seal**
  - See face seal .
- **B**
- **B_{10}-life**
  - See rating life .
- **Babbitt metal**
  - A nonferrous bearing alloy originated by Isaac Babbitt in 1839. Currently, the term includes several tin-base alloys consisting mainly of various amounts of copper, antimony, tin, and lead. Lead-base Babbitt metals are also used (Ref 1).
- **back ring**
  - A split or multisegment ring in a circumferential seal assembly used for restricting axial leakage flow (Ref 4m).
- **back-to-back ring seal**
  - An adaptation of the simple ring seal that employs two identical elements loaded axially by a spring placed between the rings. The spring forces the elements against mating rings on either side (Ref 4).
- **backing**
  - In plain bearings, the part of the bearing to which the bearing alloy is attached, normally by a metallurgical bond (Ref 1m).
- **ball bearing**

- A rolling-element bearing in which the rolling elements are spherical.
- **ball complement**
  - The number of balls contained in a ball bearing (Ref 1).
- **ball indented bearing**
  - A bearing with surface indentations serving as lubricant reservoirs (Ref 1).
- **basic dynamic load capacity**
  - (of a bearing). The radial load that a rolling-element bearing can support for a rating life of one million revolutions (500 h at $33\frac{1}{3}$ rpm) (Ref 1m). See also basic load rating .
- **basic load rating**
  - (*C*). The radial load that a ball bearing can withstand for one million revolutions of the inner ring. *Note:* The value of the basic load rating depends on bearing type, bearing geometry, accuracy of fabrication, and bearing material (Ref 6m). See also basic dynamic load capacity and dynamic load .
- **basic static load rating**
  - (of a bearing). A load that, if exceeded on a nonrotating rolling-element bearing, produces a total permanent deformation of rolling element and race at the most heavily stressed contact point of 0.0001 times the ball or roller diameter of greater (Ref 1m).
- **bath lubrication**
  - See flood lubrication .
- **bearing**
  - A support or guide by means of which a moving part is located with respect to other parts of a mechanism (Ref 1).
- **bearing area**
  - (1) The projected bearing or load-carrying area when viewed in the direction of the load (Ref 1). *Note:* Sometimes used as a synonym for real area of contact (this usage is not recommended). (2) The sum of the horizontal intercepts of a surface profile at a given level (Ref 1).
- **bearing bronzes**
  - Bronzes used for bearing applications. Two common types of bearing bronzes are copper-base alloys containing 5 to 20 wt% tin and a small amount of phosphorus (phosphor bronzes) and copper-base alloys containing up to 10 wt% tin and up to 30 wt% lead (leaded bronzes) (Ref 1m).
- **bearing characteristic number**
  - A dimensionless number that is used to evaluate the operating conditions of plain bearings (Ref 1). See also capacity number and Sommerfeld number .
- **bearing fraction**
  - The ratio of the bearing area to a reference length (Ref 1).
- **Beilby layer**
  - The altered surface layer formed on a crystalline solid during a wear process or by mechanical polishing that, according to a theory by Sir George Beilby, consists of a material in an amorphous state. *Note:* It is now generally accepted that an amorphous layer is not formed at a surface except possibly in some special cases. Depending on the polishing conditions, the surface layer may be a complex mixture of embedded polishing compound, reaction products, and plastically deformed and fractured surface material. In the context of tribology, the damaged or otherwise altered surface layer resulting from wear is often referred to as the Beilby layer without any implication as to its actual structure or nature (Ref 2).
- **bellows seal**
  - A type of mechanical seal that utilizes a bellows for providing secondary sealing (Ref 4).
- **bidirectional seal**
  - A seal that is designated to seal equally well when the pressure is applied from either direction (Ref 4).
- **big-end bearing**
  - A bearing at the larger (crankshaft) end of a connecting rod in an engine. Also known as bottom-end bearing , crankpin bearing , and large-end bearing (Ref 1m). See also little-end bearing .
- **bimetal bearing**
  - A bearing consisting of two layers. *Note:* Bimetal bearings are usually made with a layer of bearing alloy on a bronze or steel backing (Ref 1).
- **birotational seal**

- A seal that is designed for applications in which a shaft rotation is in either direction (Ref 4).
- **Bingham solid**
  - An idealized form of solid that begins to flow appreciably only when a certain stress, called the yield stress or yield point, has been exceeded. The solid subsequently flows at a rate proportional to the difference between the applied stress and this yield stress. *Note:* Many greases can be regarded as Bingham solids (Ref 1).
- **bleeding**
  - Separation of oil (or other fluid) from a grease (Ref 1).
- **blending**
  - The process of mixing mineral oils to obtain the desired consistency. *Note:* Blending should be contrasted with compounding, which utilizes additives (Ref 1m).
- **block grease**
  - A grease that is sufficiently hard to retain its shape in block or stick form (Ref 1).
- **blown oil**
  - Fatty oil that is artificially thickened by blowing air through it (Ref 1).
- **body**
  - A loosely used term designating viscosity or consistency. *Note:* Use of this term is not recommended (Ref 1m).
- **bonded film lubricant**
  - See bonded solid lubricant .
- **bonded solid lubricant**
  - A solid lubricant dispersed in a continuous matrix of a binder, or attached to a surface by an adhesive material (Ref 1).
- **bone oil**
  - A fatty oil obtained by dry distillation of bones (Ref 1).
- **bore seal**
  - A device in which the outside diameter mates with a bore surface to provide sealing between the two surfaces (Ref 4m).
- **bottom-end bearing**
  - See big-end bearing .
- **boundary lubricant**
  - A lubricant suitable for use in boundary lubrication conditions. *Note:* Fatty acids and soaps are commonly used (Ref 1.)
- **boundary lubrication**
  - A condition of lubrication in which the friction and wear between two surfaces in relative motion are determined by the properties of the surfaces and by the properties of the lubricant other than bulk viscosity. *Note:* Many circumstances that at one time were referred to as boundary lubrication may be elastohydrodynamic (Ref 1). See also elastohydrodynamic lubrication , extreme-pressure lubrication , and thin-film lubrication .
- **Brale indenter**
  - A conical 120° diamond indenter with a conical tip (a 0.2 mm tip radius is typical) used in certain types of Rockwell and scratch hardness tests.



Brale indenter

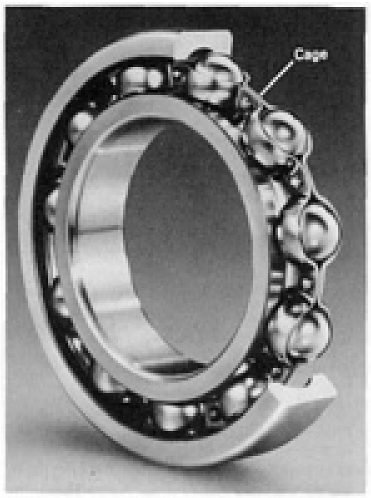
- **break-away torque**
  - See starting torque .
- **break in**
  - (verb). To operate a newly installed bearing, seal, or other tribocomponent in such a manner as to condition its surface(s) for improved functional operation. See also run in (verb).
- **break-in**
  - (noun). See running-in .
- **bright stock**
  - High-viscosity mineral oils that remain after vacuum distillation of crude oil (Ref 1).
- **Brinell hardness test**
  - Invented by Swedish engineer J.A. Brinell and gaining widespread acceptance about 1900, it is a measure of penetration hardness obtained by forcing a hard sphere of diameter (*D*) (10 mm typically) into a surface under a load (*P*) so as to produce an impression of diameter (*d*) in the surface of the testpiece. The Brinell hardness number (HB) is a measure of the average indentation pressure calculated as follows:

$$\mathrm{HB} = \frac{P}{(\pi D/2)[D - \sqrt{(D^2 - d^2)}]}$$

- **Brinelling**
  - (1) Indentation of the surface of a solid body by repeated local impact or impacts, or static overload. *Note:* Brinelling may occur especially in a rolling-element bearing (Ref 1). (2) Damage to a solid bearing surface characterized by one or more plastically formed indentations brought about by overload. *Note:* This term is often applied in the case of rolling-element bearings (Ref 2). See also false Brinelling .
- **bubbly oil**
  - Oil containing bubbles of gas (Ref 1).
- **built-up edge**
  - Material from the workpiece, especially in machining, which is stationary with respect to the tool (Ref 1). See also wedge formation .
- **burning**

- In sliding contacts, the oxidation of a surface due to local heating in an oxidizing environment (Ref 1). See also metallurgical burn .

- **burnish**
  - (1) To alter the original manufactured surface of a sliding or rolling surface to a more polished condition (Ref 1m). (2) To apply a substance to a surface by rubbing.
- **bush bearing**
  - A plain bearing in which the lining is closely fitted into the housing in the form of a bush, usually surfaced with a bearing alloy (Ref 1).
- **C**
- **cage**
  - In a bearing, a device that partly surrounds the rolling elements and travels with them, the main purpose of which is to space the rolling elements in proper relation to each other (Ref 1). See also separator .



Cage in a radial ball bearing (cutaway view)

- **capacity number**
  - (C_n). The product of the Sommerfeld number and the square of the length-to-diameter ratio of a journal bearing (Ref 7).
- **capture efficiency**
  - See collection efficiency .
- **carbonization**
  - A reduction of hydrocarbons, resulting in the formation of carbonaceous residue (Ref 4).
- **case crushing**
  - A term used to denote longitudinal gouges arising from fracture in case-hardened gears (Ref 1m).
- **catastrophic period**
  - In cavitation or liquid impingement erosion, a stage during which the erosion rate increases so dramatically that continued exposure threatens or causes gross disintegration of the exposed surface. *Note:* This stage is not inevitable; it is observed most commonly with some brittle materials. When it does occur, it may begin during any stage of the more common erosion rate-time pattern (Ref 2m).
- **catastrophic wear**
  - Sudden surface damage, deterioration, or change of shape caused by wear to such an extent that the life of the part is appreciably shortened or action is impaired (Ref 1).
- **catchment efficiency**
  - See collection efficiency .
- **cavitating disk apparatus**
  - A flow cavitation test device in which cavitating wakes are produced by holes in, or protuberances on, a disk rotating within a liquid-filled chamber. *Note:* Erosion test specimens are

attached flush with the surface of the disk at the location where the bubbles are presumed to collapse (Ref 2m).

- **cavitation**
  - The formation and collapse, within a liquid, of cavities or bubbles that contain vapor or gas or both. *Notes:* In general, cavitation originates from a decrease in the static pressure in the liquid. It is distinguished in this way from boiling, which originates from an increase in the liquid temperature. There are certain situations where it may be difficult to make a clear distinction between cavitation and boiling, and the more general definition that is given here is therefore to be preferred. In order to erode a solid surface by cavitation, it is necessary for the cavitation bubbles to collapse on or close to that surface (Ref 2).
- **cavitation cloud**
  - A collection of a large number of cavitation bubbles. *Note:* The bubbles in a cloud are small, typically less than 1 mm in cross section (Ref 2m).
- **cavitation erosion**
  - (1) Progressive loss of original material from a solid surface due to continued exposure to cavitation (Ref 2). (2) Wear of a solid body moving relatively to a liquid in a region of collapsing vapor bubbles that cause local high-impact pressures or temperatures. *Notes:* It is recommended that the term cavitation erosion should be used in preference to cavitation to describe the wear process; cavitation refers only to the formation and collapse of cavities in the liquid. Cavitation erosion may also be intensified by chemical action (Ref 1).
- **cavitation number**
  - (σ). A dimensionless number that measures the tendency for cavitation to occur in a flowing stream of liquid, and that is computed from the equation:

$$\sigma = (P_0 - P_v)/\tfrac{1}{2}\rho V_0^2$$

  - where P_v is the vapor pressure, P_o is the static pressure in the stream in an undisturbed state, V_o is the undisturbed stream velocity, and ρ is the liquid density. *Note:* The cavitation number and the net positive suction head (NPSH) are related by the following equation:

$$\text{NPSH} = (\sigma + 1)\, V_0^2/2g$$

  - where g is the acceleration due to gravity (Ref 2).
- **cavitation tunnel**
  - A flow cavitation test facility in which liquid is pumped through a pipe or tunnel, and cavitation is induced in a test section by conducting the flow through a constriction, or around an obstacle, or a combination of these (Ref 2).
- **centane number**
  - A measure of the ignition quality of a fuel or petroleum product with reference to normal centane high-ignition quality fuel with an arbitrary number of 100 (Ref 5).
- **chafing**
  - Repeated rubbing between two solid bodies that can result in surface damage and/or wear.
- **channeling**
  - (1) The tendency of a grease or viscous oil to form air channels in a bearing or gear system, resulting in an incomplete lubricant film (Ref 1). (2) The tendency of a grease to form a channel by working down in a bearing or distribution system, leaving shoulders to act as a reservoir and seal (Ref 1).
- **chatter**
  - Elastic vibrations resulting from frictional or other instability (Ref 1).
- **checking**
  - See craze cracking .
- **chemical conversion coating**
  - A coating produced by a chemical or electrochemical treatment of a metallic surface resulting in a superficial layer containing compounds of the metal. Such coatings are usually pretreatments

onto which fluid or solid lubricants may be applied, but which may themselves have lubricating properties (Ref 1).

- **chemical wear**
  - See corrosive wear .
- **chlorinated lubricant**
  - A lubricant containing a chlorine compound that reacts with a rubbing surface at elevated temperatures to protect it from sliding damage (Ref 1). See also extreme-pressure lubricant , sulfochlorinated lubricant , and sulfurized lubricant .
- **circular-step bearing**
  - A flat circular hydrostatic bearing with a central circular recess (Ref 1). See also step bearing .
- **cladding**
  - A layer of material, usually metallic, that is mechanically or metallurgically bonded to a substrate. *Note:* Cladding may be bonded to the substrate by any of several processes, such as roll-cladding and explosive forming.
- **clearance ratio**
  - In a bearing, the ratio of radial clearance to shaft radius (Ref 1).
- **cloud point**
  - The temperature at which a wax cloud first appears on cooling a mineral oil under specified conditions (Ref 1).
- **cocoa**
  - In fretting wear, a powdery form of debris, usually consisting of iron oxides, that is expelled from a ferrous metal joint near the location where fretting wear is occurring. Also known as red mud .
- **coefficient of adhesion**
  - (1) The ratio of the normal force required to separate two bodies to the normal load with which they were previously placed together. Used in studies with denuded surfaces (Ref 1). (2) In railway engineering, sometimes used to signify the coefficient of (limiting) static friction (Ref 1).
- **coefficient of friction**
  - In tribology, the dimensionless ratio of the friction force (F) between two bodies to the normal force (N) pressing these bodies together (Ref 2):

$$\mu (\text{or } f) = (F/N)$$

  - *Notes:* The friction force is often proportional to the normal force when the surfaces rub together under dry or boundary-lubricated conditions, but not when the surfaces are separated by a full fluid film, as in hydrodynamic lubrication, or by a solid-film lubricant. The friction force and normal force are not proportional for denuded surfaces or for single contacts between elastically deformed bodies. Two symbols, f and μ, are commonly used for designating the coefficient of friction, but subscripts may be added to further specify static or kinetic friction conditions (Ref 1m).
- **coefficient of wear**
  - See Archard wear law , wear coefficient , wear constant , and wear factor .
- **cold test**
  - A test in which the pour point of an oil is determined (Ref 1).
- **collar oiler**
  - A collar on a shaft that extends into the oil reservoir and carries oil into a bearing as the shaft rotates. Wipers are usually provided to direct the oil into the bearing (Ref 1).
- **collection efficiency**
  - The cross-sectional area of undisturbed fluid containing particles that will ultimately impinge on a given solid surface, divided by the projected area of the solid surface (Ref 2m). Also known as collision efficiency , capture efficiency , catchment efficiency , and impaction ratio .
- **collision efficiency**
  - See collection efficiency .
- **compatibility (frictional)**
  - In tribology, materials that exhibit good sliding behavior, including resistance to adhesive wear, are termed frictionally compatible. *Note:* Under some conditions materials that are not normally

considered compatible in the metallurgical sense (for example, silver and iron) may be very compatible in the frictional sense (Ref 1m).

- **compatibility (lubricant)**
  - In tribology, a measure of the degree to which lubricants or lubricant components can be mixed without harmful effects such as formation of deposits (Ref 1).
- **compatibility (metallurgical)**
  - A measure of the extent to which materials are mutually soluble in the solid state (Ref 1).
- **compensation**
  - In a hydrostatic bearing, the use of flow resistances between the supply and the entry to the bearing (Ref 1).
- **composite bearing material**
  - A solid material composed of a continuous or particulate solid lubricant dispersed throughout a load-bearing matrix to provide continuous replenishment of solid lubricant films as wear occurs, and effective heat transfer from the friction surface (Ref 1). *Note:* Composite bearing materials need not only be self-lubricating, but can also be used for low wear, good frictional heat dissipation, or controlled frictional response (for example, brake materials). Examples of composite materials are metal-matrix composites, laminated composites, filament-wound composites, and fiber-mineral-metal-filled composites (Ref 8).
- **compressibility number**
  - A dimensionless group used in gas bearing calculations. *Note:* For thrust bearings, the compressibility number relates viscosity (η), velocity (v), pressure (P_a), and clearance (h_o) as follows:

$$\frac{\eta v}{P_a h_0^2}$$

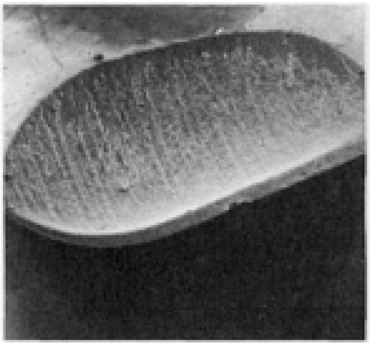
  - For journal bearings, the compressibility number relates to viscosity (η), rotational speed (U), pressure (P_o), radius (r), and clearance (c) as follows (Ref 1):

$$\frac{\eta U r}{P_0 c^2}$$

- **compression set**
  - In seals, the difference between the thickness of a gasket of static seal before the seal is compressed and after it is released from compression. *Note:* Compression set is normally expressed as a percentage of the total compression (Ref 4m).
- **cone resistance value**
  - (CRV). A measure of the yield stress of a grease, obtained by static indentation with a cone. *Note:* The equilibrium depth of penetration is measured, not the penetration in a given time, which is the penetration value (Ref 1). See also penetration (of a grease) .
- **conformability**
  - In tribology, that quality of a plain bearing material that allows it to adjust itself to shaft deflections and minor misalignments by deformation or by wearing away of bearing material without producing operating difficulties (Ref 1).
- **conformal surfaces**
  - Surfaces whose centers of curvature are on the same side of the interface. *Note:* In wear testing, it refers to the case where the curvature of both specimens matches such that the nominal contact area during the testing remains approximately constant (Ref 1m). Contrast with counterformal surfaces .
- **consistency**
  - An imprecise measure of the degree to which a grease resists deformation under the application of a force (Ref 1). See also cone resistance value and penetration hardness number .
- **constriction resistance**

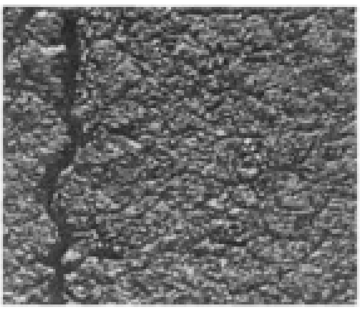
- In electrical contact theory, the resistance that arises from the constriction of current flow lines in order to pass through small areas of contact (a-spots) at the interface of two contacting bodies (Ref 1m). See also a-spot , contact resistance , and film resistance .

- **contact angle**
  - In lubrication, the angle at which the surface of a liquid drop meets the surface of a solid on which it is placed (Ref 1).
- **contact angle**
  - (in a bearing). See angle of contact .
- **contact lubrication**
  - A little-used term relating to the conditions of lubrication obtained with solid lubricant powders rubbed into a surface. *Note:* It is recommended that this term should not be used (Ref 1).
- **contact resistance**
  - The electrical resistance between two contacting bodies, which is the sum of the constriction resistance and the film resistance (Ref 1).
- **contact stress**
  - Stress that results near the surfaces of two contacting solid bodies when they are placed against one another under a nonzero normal force. *Note:* This term is not sufficiently precise because it does not indicate the type of stress, although common usage usually implies elastic or Hertz stress.
- **contacting ring seal**
  - A type of circumferential seal that utilizes a ring spring-loaded radially against a shaft. The ring is either gapped or segmented, in order to have radial flexibility. The seal has overlapping joints for blocking leakage at the gaps. An axial spring load seats the ring against the wall of its containing cartridge (Ref 4m).
- **conversion coating**
  - See chemical conversion coating .
- **corrosive wear**
  - A wear process in which chemical or electrochemical reaction with the environment predominates (Ref 1). See also oxidative wear .
- **Coulomb friction**
  - A term used to indicate that the frictional force is proportional to the normal load (Ref 1).
- **Coulomb's laws**
  - See Amontons' laws .
- **counterformal surfaces**
  - Surfaces whose centers of curvature are on the opposite sides of the interface, as in rolling-element bearings or gear teeth. *Note:* In wear testing, this term is sometimes used to indicate that the test specimen surfaces are not conformal (for example, with a sphere-on-flat or flat block-on-rotating ring configuration). In such cases, the nominal contact area of at least one of the testpiece surfaces increases as the amount of wear increases (Ref 1m). Contrast with conformal surfaces .
- **cracking**
  - In lubrication technology, that process of converting unwanted long-chain hydrocarbons to shorter molecules by thermal or catalytic action (Ref 1).
- **crankpin bearing**
  - See big-end bearing .
- **crater wear**
  - The wear that occurs on the rake face of a cutting tool due to contact with the material in the chip that is sliding along that face.



Crater wear on a cemented carbide tool. 15×

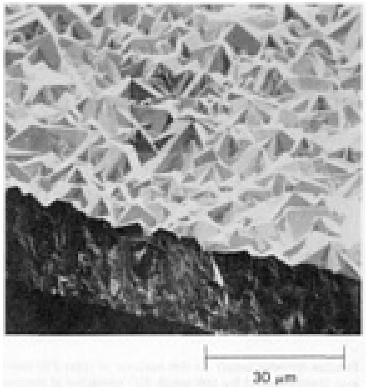
- **craze cracking**
  - Irregular surface cracking associated with thermal cycling. *Note:* This term is used more in the United Kingdom than in the United States, where the term checking is used instead (Ref 1m).



Craze cracking on the outer surface of a stainless steel tube. ~4×

- **creep**
  - (1) Time-dependent deformation. *Note:* In tribology, the deformation may occur if a static or moving contact is held under load at such conditions that one or more of the materials in contact experiences time-dependent deformation. Creep is a problem in certain polymeric bearing materials. (2) Small relative tangential displacement in a contacting area at an interface, when the remainder of the interface in the contacting area is not relatively displaced tangentially (Ref 3). Also known as microslip .
- **creep ratio**
  - In a ball bearing, the ratio of the creep velocity to sweep velocity (Ref 1m). See also creep (2).
- **crush**
  - In a split-journal bearing, the amount by which a bearing half extends above the horizontal split of the bore before it is assembled (Ref 1m). Also known as nip in the United Kingdom.
- **cumulative erosion-time curve**
  - A plot of cumulative erosion versus cumulative exposure duration, usually obtained by periodic interruption of the erosion test and weighing of the specimen. *Note:* This is the primary record of an erosion test used to derive other characteristics such as incubation period, maximum erosion rate, terminal erosion rate, and the erosion rate-time curve (Ref 2m).
- **cut**
  - In lubricant technology, a product or fraction obtained by distillation within a specified temperature range 1
- **cutting wear**
  - See abrasive wear
- **D**
- **debris**
  - See wear debris .

- **deceleration period**
  - In cavitation or impingement erosion, the stage following the acceleration period or the maximum rate period (if any), during which the erosion rate has an overall decreasing trend although fluctuations may be superimposed on it (Ref 2).
- **deformability**
  - See conformability .
- **deformation wear**
  - Sliding wear involving plastic deformation of the wearing surface. *Note:* Many forms of wear involve plastic deformation, so this term is imprecise and should not be used.
- **delamination wear**
  - A wear process in which thin layers of material are formed and removed from the wear surface. *Note:* A "delamination theory of wear" proposed by N.P. Suh in the early 1970s involves the nucleation and propagation of cracks so as to form lamellar wear particles; however, there are other ways to form lamellar wear particles (for example, by the loss of transfer layer fragments), and the term delamination wear should not be used unless further clarification of the context of usage is provided.
- **detergent additive**
  - In lubrication technology, a surface-active additive that helps to keep solid particles suspended in an oil (Ref 1m).
- **detergent oil**
  - A heavy-duty oil containing a detergent additive. *Note:* Detergent oils are used mainly in combustion engines (Ref 1).
- **detritus**
  - See wear debris .
- **diamond film**
  - A carbon-composed film, usually deposited by chemical vapor deposition or related process, that has the following three characteristics: (1) a crystalline morphology that can be visually discerned by optical microscopy, (2) a single-phase crystalline structure identifiable by x-ray and/or electron diffraction, and (3) a Raman spectrum typical for crystalline diamond. *Note:* The sharpness of the sp^3 carbon-carbon bonding peak (wave-number 1332 cm^{-1}) is often used to determine the perfection of the film (Ref 9m). See also diamondlike film .



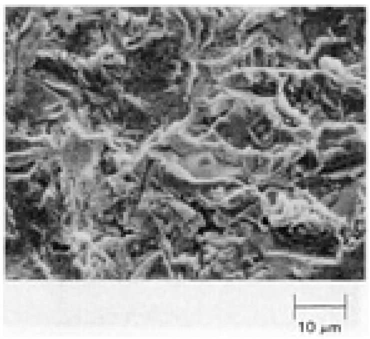
Diamond film grown on silicon. Courtesy of R. Clausing, ORNL

- **diamondlike film**
  - A hard, noncrystalline carbon film, usually grown by chemical vapor deposition or related techniques, that contains predominantly sp^2 carbon-carbon bonds. *Note:* In Raman spectroscopy, the absence of a sharp peak at wavenumber 1332 cm^{-1} and a broad peak at 1500 to 1600 cm^{-1} is characteristic of diamondlike films (Ref 9m). See also diamond film .
- **dilatant**
  - A reversible increase in viscosity with increasing shear stress (Ref 1). Compare with pseudoplastic behavior , rheopectic material , and thixotropy .
- **dirt content**

- A measure of the size and concentration of foreign particles present in a lubricant. *Note:* Dirt content is usually reported as the number of particles per cubic centimeter, for specified particle sizes (Ref 5).
- **disk machine**
  - In tribology, a testing machine for rolling or rolling/sliding contact in which two disk-shaped rollers, with parallel axes of rotation, make tangential contact on their circumferences as they move relative to one another. *Note:* One or both disks may be crowned on the rolling-contact surface. See also Amsler wear machine .
- **dispersant additive**
  - In lubrication technology, an additive capable of dispersing cold oil sludge (Ref 1m).
- **dispersant oil**
  - A heavy-duty oil containing a dispersant additive (Ref 1).
- **distributed impact test**
  - In impingement erosion testing, an apparatus or method that produces a spatial distribution of impacts by liquid or solid bodies over an exposed surface of a specimen. *Note:* Examples of such tests are those employing liquid sprays or simulated rainfields. If the impacts are distributed uniformly over the surface, the term uniformly distributed impact test may be used (Ref 2).
- **DN value**
  - The product of bearing bore diameter in millimeters and speed in revolutions per minute (Ref 5).
- **drawing compound**
  - In metalworking, a lubricant having extreme-pressure properties (Ref 1m). See also extreme-pressure lubricant .
- **drip feed (drop feed) lubrication**
  - A system of lubrication in which the lubricant is supplied to the bearing surfaces in the form of drops at regular intervals (Ref 1m).
- **drop (dropping) point**
  - The temperature at which a drop falls from a grease through a specified orifice. *Note:* The drop point does not necessarily represent the maximum operating temperature of a grease (Ref 1m).
- **droplet erosion**
  - Erosive wear caused by the impingement of liquid droplets on a solid surface. See also erosion (erosive) wear .
- **dry-film lubrication**
  - Lubrication that involves the application of a thin film of solid lubricant to the surface or surfaces to be lubricated.
- **dry friction**
  - Friction that occurs between two bodies in the absence of lubrication. *Note:* This term is inaccurate because it historically implies that there is no intentionally applied lubrication, when in fact solid lubrication conditions can be considered "dry". Therefore, this term should not be used.
- **dry lubricant**
  - See solid lubricant . *Note:* It has been suggested that this term should not be used as a synonym for solid lubricant because some solid lubricants perform better in the presence of moisture (for example, graphite). Instead, it should refer to lubricants that function in low-moisture environments (Ref 10).
- **dry-running**
  - In seals, running without liquid present at the seal surface (Ref 4).
- **dry sand-rubber wheel test**
  - In wear testing, a term used to describe a standard abrasive wear testing method in which a stream of dry quartz sand is passed between a rotating rubber wheel and a stationary test coupon that is held against it under specified normal force. Further information can be found in ASTM Standard Test Method G 65.
- **dry sliding water**
  - Sliding wear in which there is no intentional lubricant or moisture introduced into the contact area. *Note:* Species such as oxides and other films, or adsorbed species that are initially present or that form during sliding contact, may lubricate the contact; however, the contact is often referred to as being "unlubricated". This imprecise term is not usually used when solid lubricants are introduced even though they could be considered to be dry. See also unlubricated sliding .
- **duplex alloys**

- Bearing alloys consisting of two phases, one much softer than the other (Ref 1).
- **durometer reading**
  - An index that is used for ranking the relative hardness of elastomers (Ref 4). *Note:* The durometer hardness test involves forcing a 30° tapered indenter into the surface of the specimen using calibrated loading springs. A dial gage indicates the depth of penetration in durometer numbers, which are directly proportional to the load on the spring.
- **dusting**
  - A phenomenon, usually affecting carbon-base electrical motor brushes or other current-carrying contacts, wherein at low relative humidity or high applied current density, a powdery "dust" is produced during operation.
- **duty**
  - The specification giving the load, ambient temperature, and speed under which surfaces are required to move (Ref 1).
- **duty parameter**
  - See capacity number .
- **dynamic friction**
  - See kinetic friction .
- **dynamic load**
  - An imposed force that is in motion, that is, one that may vary in magnitude, sense, and direction (Ref 11).
- **dynamic seal**
  - A seal that has rotating, oscillating, or reciprocating motion between its components, as opposed to stationary-type seal such as a gasket (Ref 4).
- **dynamic viscosity**
  - See viscosity .
- **E**
- **eccentricity**
  - In journal bearings, the radial displacement of the journal center from the center of the bearing liner (Ref 1m).
- **eccentricity ratio**
  - In a bearing, the ratio of the eccentricity to the radial clearance (Ref 1).
- **effective leakage area**
  - The orifice flow area that will result in the same calculated flow for a given pressure drop as is measured for the seal in question. *Note:* This concept is useful when comparing the leakage performance of seals of different sizes and designs, and of seals operating under different conditions (Ref 4m).
- **elastic compliance**
  - A condition under which two bodies in contact, which are subjected to a force, undergo small elastic displacement without slip (Ref 1).
- **elastohydrodynamic lubrication**
  - A condition of lubrication in which the friction and film thickness between two bodies in relative motion are determined by the elastic properties of the bodies, in combination with the viscous properties of the lubricant at the prevailing pressure, temperature, and rate of shear (Ref 1). See also boundary lubrication , plastohydrodynamic lubrication , and thin-film lubrication .
- **electrical pitting**
  - The formation of surface cavities by removal of metal as a result of an electrical discharge across an interface (Ref 1).
- **electrocorrosive wear**
  - Wear to a solid surface which is accelerated by the presence of a corrosion-inducing electrical potential across the contact interface. *Note:* This process is usually associated with wear in the presence of a liquid electrolyte in the interface. However, moisture from the air can also facilitate this type of wear when a galvanic wear couple exists and the contacting materials are sufficiently reactive.
- **elliptical bearing**
  - See lemon bearing (elliptical bearing) .
- **embeddability**
  - The ability of a bearing material to embed harmful foreign particles and reduce their tendency to cause scoring or abrasion (Ref 1).

- **emulsion**
  - A dispersion of globules of one liquid in a second immiscible liquid (for example, oil in water). *Notes:* Lubricant emulsions are usually stabilized by a suitable agent. The stability of an emulsion increases with decreasing diameter of the droplets. Typical lubricant emulsions are prepared from oil and water (Ref 1).
- **emulsion inversion**
  - An emulsion is said to invert when, for example, a water-in-oil emulsion changes to an oil-in-water emulsion (Ref 1).
- **end seal**
  - See face seal .
- **engine oil**
  - An oil used to lubricate an internal combustion engine (Ref 1).
- **Engler viscosity**
  - A commercial measure of viscosity expressed as the ratio between the time in seconds required for 200 cm^3 of a fluid to flow through the orifice of an Engler viscometer at a given temperature under specified conditions and the time required for 200 cm^3 of distilled water at 20 °C to flow through the orifice under the same conditions. *Note:* It is recommended that standard viscosity units should be used (Ref 1).
- **entraining velocity**
  - The velocity of a liquid at which bubbles of gas are carried along in the stream (Ref 1).
- **EP lubricant**
  - See extreme-pressure lubricant .
- **equivalent radial load**
  - The level of constant radial load on a rolling-element bearing that, when the bearing is stationary with respect to the outer race, will produce the same rating life as a given combination of radial and thrust loads under the same conditions of operation (Ref 1).
- **erosion (erosive wear)**
  - (1) Loss of material from a solid surface due to relative motion in contact with a fluid that contains solid particles. *Note:* Erosion in which the relative motion of particles is nearly parallel to the solid surface is called abrasive erosion . Erosion in which the relative motion of the solid particles is nearly normal to the solid surface is called impingement erosion or impact erosion (Ref 1). (2) Progressive loss of original material from a solid surface due to mechanical interaction between that surface and a fluid, a multicomponent fluid, and impinging liquid, or solid particles. *Note:* Because of the broad scope of this term, it is recommended that it normally be qualified to indicate the relevant mechanism or context, for example, impingement erosion , abrasive erosion , and so forth (Ref 2). See also cavitation erosion and electrical pitting .



Erosion (erosive wear) on the surface of type 310 stainless steel eroded by 100 mesh SiC particles at normal incidence and 40 m/s impingement velocity. Courtesy of L.K. Ives, NIST

- **erosion-corrosion**
  - A conjoint action involving corrosion and erosion in the presence of a corrosive substance (Ref 2).
- **erosion rate**

- Any determination of the rate of loss of material (erosion) with exposure duration. *Note:* In certain contexts (for example, ASTM erosion tests), it is given by the slope of the cumulative erosion-time curve (Ref 2m).

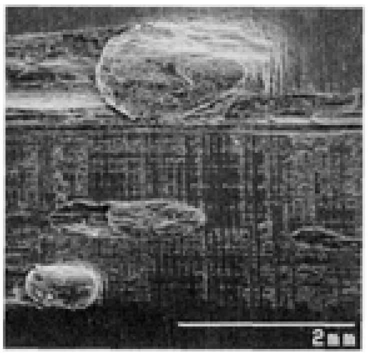
- **erosion rate-time curve**
  - A plot of instantaneous erosion rate versus exposure duration, usually obtained by numerical or graphical differentiation of the cumulative erosion-time curve (Ref 2).
- **erosivity**
  - The characteristic of a collection of particles, liquid stream, or a slurry that expresses its tendency to cause erosive wear when forced against a solid surface under relative motion.
- **excrescence**
  - A term used by some Russian tribologists (in English) describing micro-extrusions on friction surfaces that lead to localized welding. *Note:* This term is not recommended (Ref 1).
- **externally pressurized seal**
  - A seal that operates on a thin film at the interface with the mating surface. *Note:* The film is formed by high-pressure fluid that is brought to the interface at some mid-dam location and that is at a pressure equal to, or higher than, the up-stream seal pressure (Ref 4m). See also hydrostatic seal .
- **extreme-pressure lubricant**
  - A lubricant that imparts increased load-carrying capacity to rubbing surfaces under severe operating conditions. *Note:* Extreme-pressure lubricants usually contain sulfur, halogens, or phosphorus. The term antiscuffing lubricant has been suggested as a replacement for extreme-pressure lubricant (Ref 1m).
- **extreme-pressure lubrication**
  - A condition of lubrication in which the friction and wear between two surfaces in relative motion depend upon the reaction of the lubricant with a rubbing surface at elevated temperature (Ref 1).
- **F**
- **face pressure**
  - The face load divided by the contacting area of the sealing lip. *Notes:* The face load is the sum of the pneumatic or hydraulic force and the spring force. For lip seals and packings, the face load also includes the interference load (Ref 4m).
- **face seal**
  - A device that prevents the leakage of fluids along rotating shafts. *Note:* Sealing is accomplished by a stationary primary-seal ring bearing against the face of a mating ring mounted on a shaft. Axial pressure maintains the contact between seal ring and mating ring (Ref 4m).
- **false Brinelling**
  - (1) Damage to a solid bearing surface characterized by indentations not caused by plastic deformation resulting from overload, but thought to be due to other causes such as fretting corrosion (Ref 2). (2) Local spots appearing when the protective film on a metal is broken continually by repeated impacts, usually in the presence of corrosive agents. *Notes:* The appearance is generally similar to that produced by Brinelling but corrosion products are usually visible. It may result from fretting corrosion. This term should be avoided when a more precise description is possible (Ref 1). False Brinelling (race fretting) can be distinguished from true Brinelling because in false Brinelling, surface material is removed so that original finishing marks are removed. The borders of a false Brinell mark are sharply defined, whereas a dent caused by a rolling element does not have sharp edges and the finishing marks are visible in the bottom of the dent (Ref 8).
- **fat**
  - An organic ester, the product of a reaction between a fatty acid and glycerol. *Note:* Fat can be of animal or vegetable materials or can be made synthetically (Ref 1).
- **fatigue wear**
  - (1) Removal of particles detached by fatigue arising from cyclic stress variations (Ref 1). (2) Wear of a solid surface caused by fracture arising from material fatigue (Ref 2). See also spalling .
- **fatty acid**
  - An organic acid of aliphatic structure originally derived from fats and fatty oils (Ref 1).
- **fatty oil**
  - A fat (glycerol ester) that is liquid at room temperature (Ref 1).
- **feedability**

- The ability of a grease to flow to the suction of a pump (Ref 1).
- **ferrograph**
  - An instrument used to determine the size distribution of wear particles in lubricating oils of mechanical systems (Ref 5). *Note:* The technique relies on the debris being capable of being attracted to a magnet.
- **fiber grease**
  - A type of grease having a pronounced fibrous structure (Ref 1).
- **file hardness**
  - Hardness as determined by the use of a file of standardized hardness on the assumption that a material that cannot be cut with the file is as hard as, or harder than, the file. *Note:* Files covering a range of hardnesses may be employed (Ref 5m).
- **filler**
  - In lubrication, a substance such as lime, talc, mica, and other powders, added to a grease to increase its consistency or to an oil to increase its viscosity (Ref 1).
- **film resistance**
  - The electrical resistance that results from films at contacting surfaces, such as oxides and contaminants, that prevent pure metallic contact (Ref 1). See also constriction resistance and contact resistance .
- **film strength**
  - An imprecise term denoting ability of a surface film to resist rupture by the penetration of asperities during sliding or rolling. *Note:* A high film strength is primarily inferred from a high load-carrying capacity and is seldom directly measured. It is recommended that this term should not be used (Ref 1).
- **film thickness**
  - In a dynamic seal, the distance separating the two surfaces that form the primary seal (Ref 4m).
- **fire point**
  - The temperature at which a material will continue to burn for at least 5 s without the benefit of an outside flame (Ref 5).
- **fixed-land bearing**
  - See fixed-pad bearing .
- **fixed oil**
  - An imprecise term denoting an oil that is difficult to distill without decomposition (Ref 1).
- **fixed-pad bearing**
  - An axial- or radial-load bearing equipped with fixed pads, the surfaces of which are contoured to promote hydrodynamic lubrication (Ref 1).
- **flaking**
  - (1) The removal of material from a surface in the form of flakes or scalelike particles. (2) A form of pitting resulting from fatigue (Ref 12m). See also spalling .
- **flank wear**
  - The loss of relief on the flank of a cutting tool behind the cutting edge due to rubbing contact between the work and the tool during cutting, measured in terms of linear dimension behind the original cutting edge (Ref 5m).
- **flash point**
  - The lowest temperature at which the vapor of a lubricant can be ignited under specified conditions (Ref 1).
- **flash temperature**
  - The maximum local temperature generated at some point in a sliding contact. *Notes:* The flash temperature occurs at areas of real contact due to the frictional heat dissipated at these areas. The duration of the flash temperature is often of the order of a microsecond. The term flash temperature may also mean the average temperature over a restricted contact area (for example, between gear teeth) (Ref 1m).
- **flexible hinge**
  - See flexure pivot bearing .
- **flexure pivot bearing**
  - A type of bearing guiding the moving parts by flexure of an elastic member or members rather than by rolling or sliding. *Note:* Only limited movement is possible with a flexure pivot (Ref 1).
- **floating bearing**
  - A bearing designed or mounted to permit axial displacement between shaft and housing (Ref 1).

- **floating-ring bearing**
  - A type of journal bearing that includes a thin ring between the journal and the bearing. The ring floats and rotates at a fraction of the journal rotational speed (Ref 1m).
- **flock point**
  - A measure of the tendency of a lubricant to precipitate wax or other solids from solution. *Note:* Depending on the test used, the flock point is the temperature required for precipitation, or the time required at a given temperature for precipitation (Ref 5).
- **flood lubrication**
  - A system of lubrication in which the lubricant is supplied in a continuous stream at low pressure and subsequently drains away. Also known as bath lubrication (Ref 1m).
- **flow cavitation**
  - Cavitation caused by a decrease in static pressure induced by changes in the velocity of a flowing liquid. *Note:* Typically this may be caused by flow around an obstacle or through a constriction, or relative to a blade or foil (Ref 2m).
- **fluid bearing**
  - See hydrostatic bearing .
- **fluid erosion**
  - See liquid impingement erosion .
- **fluid friction**
  - Frictional resistance due to the viscous or rheological flow of fluids (Ref 1m).
- **fluidity**
  - The reciprocal of viscosity . *Note:* In the centimeter-gram-second (cgs) system, the unit of fluidity is the rhe, which has units of grams per centimeter second (Ref 5).
- **fluted bearing**
  - A sleeve bearing with oil grooves generally in an axial direction (Ref 1).
- **fluting**
  - (1) In bearings, a form of pitting in which the pits occur in a regular pattern so as to form grooves (Ref 1m). *Notes:* Ridges may occur with or without burnt craters. The general cause involves vibration together with excessive wear or excessive load (Ref 12m). (2) Electric discharge pitting in a rolling-contact bearing subject to vibration (Ref 8).
- **foam inhibitor**
  - A surface-active chemical compound used in minute quantities to prevent or reduce foaming. *Note:* Silicone fluids are frequently used as foam inhibitors (Ref 1).
- **foaming**
  - In tribology, the production and coalescence of gas bubbles on a liquid lubricant surface (Ref 1).
- **foil bearing**
  - A bearing in which the housing is replaced by a flexible foil held under tension against a partition of the journal periphery, lubricant being retained between the journal and the foil (Ref 1).
- **follower plate**
  - A plate fitted to the top surface of a grease dispenser (Ref 1).
- **force-feed lubrication**
  - See pressure lubrication .
- **free rolling**
  - Rolling in which no traction is deliberately applied between a rolling element and another surface (Ref 1).
- **free spread**
  - See nip .
- **fretting**
  - (1) Wear phenomena occurring between two surfaces having oscillatory relative motion of small amplitude. *Note:* Fretting is a term frequently used to include fretting corrosion . This usage is not recommended (Ref 1). (2) Small-amplitude oscillatory motion, usually tangential, between two solid surfaces in contact. *Note:* Here the term fretting refers only to the nature of the motion without reference to wear, corrosion, or other damage that may ensue. The term fretting is often used to denote fretting corrosion and other forms of fretting wear . Usage in this sense is discouraged due to the ambiguity that may arise (Ref 2).
- **fretting corrosion**
  - A form of fretting in which chemical reaction predominates. *Notes:* Fretting corrosion is often characterized by the removal of particles and subsequent formation of oxides, which are often

abrasive and so increase the wear. Fretting corrosion can involve other chemical reaction products, which may not be abrasive (Ref 1).

- **fretting fatigue**
  - The progressive damage to a solid surface that arises from fretting. *Note:* If particles of wear debris are produced, then the term fretting wear may be applied.
- **fretting wear**
  - Wear arising as a result of fretting (2) (Ref 2).
- **friction**
  - The resisting force tangential to the common boundary between two bodies when, under the action of an external force, one body moves or tends to move relative to the surface of the other. *Note:* The term friction is also used, incorrectly, to denote coefficient of friction . It is vague and imprecise unless accompanied by the appropriate modifiers, such as dry friction or kinetic friction (Ref 1m). See also static coefficient of friction .
- **friction coefficient**
  - See coefficient of friction .
- **friction force**
  - The resisting force tangential to the interface between two bodies when, under the action of an external force, one body moves or tends to move relative to the other (Ref 2).
- **friction polymer**
  - An amorphous organic deposit that is produced when certain metals are rubbed together in the presence of organic liquids or gases (Ref 1). *Note:* Friction polymer often forms on moving electrical contacts exposed to industrial environments. The varnishlike film will attenuate or modify transmitted signals (Ref 8).
- **frosting**
  - A form of ball bearing groove damage, appearing as a frosted area, suggestive that surface distress has occurred.
- **full-film lubrication**
  - A type of lubrication wherein the solid surfaces are separated completely by an elastohydrodynamic fluid film. See also elastohydrodynamic lubrication .
- **full journal bearing**
  - A journal bearing that surrounds the journal by a full 360 ° (Ref 1m).
- **G**
- **galling**
  - A severe form of scuffing associated with gross damage to the surfaces or failure (Ref 1). *Note:* Galling has been used in many ways in tribology; therefore, each time it is encountered its meaning must be ascertained from the specific context of the usage. See also scoring and scuffing .



Galling on the surface of type 316 stainless steel rubbed against a button of the same material at a load of ~1000 N. Courtesy of L.K. Ives, NIST

- **gas bearing**
  - A journal or thrust bearing lubricated with gas.
- **gas lubrication**

- A system of lubrication in which the shape and relative motion of the sliding surfaces cause the formation of a gas film having sufficient pressure to separate the surfaces (Ref 1). See also pressurized gas lubrication .

- **gasket**
  - A device, usually made of a deformable material, that is used between two relatively static surfaces to prevent leakage (Ref 4m).
- **gelling agent**
  - See thickener .
- **glaze**
  - In tribology, a ceramic or other hard, smooth surface film produced by sliding (Ref 1).
- **gouging abrasion**
  - A form of high-stress abrasion in which easily observable grooves or gouges are created on the surface (Ref 3m). See also abrasion , abrasive wear , high-stress abrasion , and low-stress abrasion .
- **grease**
  - A lubricant composed of an oil thickened with a soap or other thickener to a semi-solid or solid consistency. *Notes:* A lime-base grease is prepared from lubricating oil and calcium soap. Sodium-, barium-, lithium-, and aluminum-base greases are also used. Greases may contain various additives. The liquid phase may also be a synthetic fluid (Ref 1).
- **guide bearing**
  - A bearing used for positioning a slide, or for axial alignment of a long rotating shaft (Ref 1).
- **gum**
  - In lubrication, a rubberlike, sticky deposit, black or dark brown in color, that results from the oxidation and/or polymerization of fuels and lubricating oils. *Note:* Harder deposit are described as lacquers or varnishes (Ref 1).
- **gutterway**
  - In bearings, a special type of spreader in which the groove is adjacent to the joint (Ref 1m).
- **H**
- **hair grease**
  - A grease containing horse hair or wool fiber (Ref 1).
- **half journal bearing**
  - A journal bearing extending 180° around a journal (Ref 1).
- **hard face**
  - A seal facing of high hardness that is applied to a softer material, such as by flame spraying, plasma spraying, electroplating, nitriding, carburizing, or welding (Ref 4m).
- **hardness**
  - The characteristic of a solid material that expresses its resistance to penetration or abrasion by other bodies. *Note:* Taken by itself, the term hardness is virtually meaningless. To have meaning, it must be expressed either as a relative quantity or as a numerical quantity obtained under specifically described conditions of a hardness testing procedure. Hardness is not a basic property of a solid because it depends on the method of its measurement and the scale used to express it. See also Brinell hardness test , indentation hardness , Knoop (microindentation) hardness number , microindentation hardness number , Rockwell hardness number , and Vickers (microindentation) hardness number .
- **Harrison number**
  - See compressibility number .
- **heat checking**
  - A process in which fine cracks are formed on the surface of a body in sliding contact due to the buildup of excessive frictional heat.
- **heavy-duty oil**
  - An oil that is stable against oxidation, protects bearings from corrosion, and has detergent and dispersant properties. *Note:* Heavy-duty oils are suitable for use in gasoline and diesel engines (Ref 1m).
- **herringbone bearing**
  - Any plain, sleeve, or thrust bearing with herringbone-shaped oil grooves (Ref 1).
- **Hershey number**
  - A dimensionless number that is used to evaluate the performance of bearings. It relates the load per unit width (P), surface velocity (U), and dynamic viscosity (η) as follows:

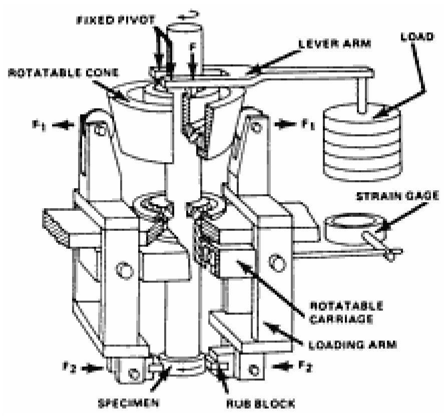
$$\frac{P}{\eta U}$$

- *Note:* This number is normally written as:

$$\frac{ZN}{p}$$

- in which *Z* denotes the dynamic viscosity, *N* the frequency of rotation, and *p* the pressure (Ref 1). See also Ocvirk number , Sommerfeld number , and Stribeck curve .

- **Hertzian contact area**
  - (1) The contact area (also, diameter or radius of contact) between two bodies calculated according to Hertz's equations of elastic deformation (Ref 1m). (2) The apparent area of contact between two nonconforming solid bodies pressed against each other, as calculated from Hertz's equations of elastic deformation (Ref 2).
- **Hertzian contact pressure**
  - (1) The pressure at a contact between two solid bodies calculated according to Hertz's equations of elastic deformation (Ref 1). (2) The magnitude of the pressure at any specified location in a Hertzian contact area, as calculated from Hertz's equations of elastic deformation (Ref 2).
- **high-stress abrasion**
  - A form of abrasion in which relatively large cutting forces are imposed on the particles or protuberances causing the abrasion, and that produces significant cutting and deformation of the wearing surface. *Note:* In metals, high-stress abrasion can result in significant surface strain hardening. This form of abrasion is common in mining and agricultural equipment, and in highly loaded bearings where hard particles are trapped between mating surfaces (Ref 3m). See also low-stress abrasion .
- **highly deformed layer**
  - In tribology, a layer of severely plastically deformed material that results from the shear stresses imposed on that region during sliding contact. See also Beilby layer and white layer .
- **Hohman A-6 wear machine**
  - A widely used type of wear and friction testing machine in which a rotating ring specimen is squeezed between two diametrically opposed rub blocks. *Note:* This design is said to eliminate shaft flexure such as that found in other machines whose load application from the rub block to the ring is from one side only. Block geometry can be changed from flat to conforming or V-block. This type of machine is designed for use with either lubricated or unlubricated specimens.



Hohman A-6 wear machine testing arrangement. Courtesy of Hohman Plating and Manufacturing Company

- **horseshoe thrust bearing**
    - A tilting-pad thrust bearing in which the top pads are omitted, making an incomplete annulus (Ref 1).
- **hydraulic fluid**
    - A fluid used for transmission of hydraulic pressure or action, not necessarily involving lubricant properties. *Note:* Hydraulic fluids can be based on oil, water, or synthetic (fire-resistant) liquids (Ref 1).
- **hydrodynamic lubrication**
    - A system of lubrication in which the shape and relative motion of the sliding surfaces causes the formation of a fluid film that has sufficient pressure to separate the surfaces (Ref 1). See also elastohydrodynamic lubrication and gas lubrication .
- **hydrodynamic seal**
    - A seal that has special geometric features on one of the mating faces. These features are designed to produce interfacial lift, which arises solely from the relative motion between the stationary and rotating portions of the seal (Ref 4).
- **hydrostatic bearing**
    - A bearing in which the solid bodies are separated and supported by a hydrostatic pressure, applied by an external source, to a compressible or incompressible fluid interposed between those bodies.
- **hydrostatic lubrication**
    - A system of lubrication in which the lubricant is supplied under sufficient external pressure to separate the opposing surfaces by a fluid film (Ref 1). See also pressurized gas lubrication .
- **hydrostatic seal**
    - A seal incorporating features that maintain an interfacial film thickness by means of pressure. The pressure is provided either by an external source or by the pressure differential across the seal. The interfacial pressure profile of a seal face is normally speed-dependent; the interfacial pressure profile of the hydrostatic seal is not speed-dependent (Ref 4).
- **hypoid gear lubricant (hypoid oil)**
    - A gear lubricant with extreme-pressure characteristics used in hypoid gears (Ref 1).
- **I**
- **impact velocity**
    - The relative velocity between the surface of a solid body and an impacting liquid or solid particle. *Note:* To describe this velocity completely, it is necessary to specify the direction of motion of the particle relative to the solid surface in addition to the magnitude of the velocity. The following related terms are also in use: (1) absolute impact velocity--the magnitude of the impact velocity, and (2) normal impact velocity--the component of the impact velocity that is perpendicular to the surface of the test solid at the point of impact (Ref 2).
- **impact wear**
    - Wear of a solid surface resulting from repeated collisions between that surface and another solid body. *Note:* The term erosion (erosive) wear is preferred in the case of multiple impacts and when the impacting body or bodies are very small relative to the surface being impacted.
- **impaction ratio**
    - See collection efficiency .
- **impingement**
    - A process resulting in a continuing succession of impacts between liquid or solid particles and a solid surface. *Note:* In preferred usage, impingement also connotes that the impacting particles are smaller than the solid surface, and that the impacts are distributed over the surface or a portion of the surface. If all impacts are superimposed on the same point or zone, then the term repeated impact is preferred (Ref 2m).
- **impingement corrosion**
    - A form of erosion-corrosion generally associated with the impingement of a high-velocity flowing liquid containing air bubbles against a solid surface (Ref 2).
- **impingement erosion**
    - Loss of material from a solid surface due to impingement .
- **impingement umbrella**
    - The partial screening of the surface of a solid specimen subjected to solid impingement that sometimes occurs when some of the solid particles rebound from the surface and impede the motion of other impinging particles (Ref 2).
- **incubation period**

- (1) In cavitation and impingement erosion, the initial stage of the erosion rate-time pattern during which the erosion rate is zero or negligible compared to later stages. (2) In cavitation and impingement erosion, the exposure duration associated with the initial stage of the erosion rate-time pattern during which the erosion rate is zero or negligible compared to later stages. *Note:* Quantitatively, incubation period is sometimes defined as the intercept on the time or exposure axis, of a straight-line extension of the maximum slope portion of the cumulative erosion-time curve (Ref 2m). See also cavitation erosion and impingement erosion .

- **indentation hardness**
  - Resistance of a solid surface to the penetration of a second, usually harder, body under prescribed conditions. *Note:* Numerical values used to express indentation hardness are not absolute physical quantities, but depend on the hardness scale used to express hardness. See also Brinell hardness test , Knoop (microindentation) hardness number , microindentation hardness number , nanohardness test , Rockwell hardness number , and Vickers (microindentation) hardness number .
- **indenter**
  - In hardness testing, a solid body of prescribed geometry, usually chosen for its high hardness, that is used to determine the resistance of a solid surface to penetration.
- **initial pitting**
  - Surface fatigue occurring during the early stages of gear operation, associated with the removal of highly stressed local areas and running-in (Ref 1).
- **instantaneous erosion rate**
  - The slope of a tangent to the cumulative erosion-time curve at a specified point on that curve (Ref 2).
- **interval erosion rate**
  - The slope of a line connecting two specified points on the cumulative erosion-time curve (Ref 2).
- **IRG transition diagram**
  - Developed by the International Research Group (IRG) on Wear of Engineering Materials of the Organization of Economic Cooperation and Development (OECD), it is a plot of normal force in newtons (ordinate) versus sliding velocity in meters per second (abscissa) wherein boundaries identify three distinct regions of varying lubricant effectiveness.
- **J**
- **jewel bearing**
  - A bearing made of diamond, sapphire, or a hard substitute metal (Ref 1).
- **journal**
  - The part of a shaft or axle that rotates or oscillates relative to a radial bearing. *Note:* A journal is part of a larger unit, for example, a crankshaft or lineshaft, and it is preferred that the term shaft be kept for the whole unit (Ref 1).
- **journal bearing**
  - A sliding-type bearing in which a journal rotates or oscillates relative to its housing. *Note:* A full journal bearing extends 360°, but partial bearings may extend, for example, over 180° or 120° (Ref 1).
- **judder**
  - Intermittent motion. See also spragging .
- **K**
- **kinematic viscosity**
  - See viscosity .
- **kinetic coefficient of friction**
  - The coefficient of friction under conditions of macroscopic relative motion between two bodies (Ref 2).
- **kinetic friction**
  - Friction under conditions of macroscopic relative motion between two bodies. *Note:* This term is sometimes used as a synonym for kinetic coefficient of friction ; however, it can also be used merely to indicate that the type of friction being indicated is associated with macroscopic motion rather than static conditions.
- **kinematic wear marks**
  - In ball bearings, a series of short curved marks on the surface of a bearing race due to the kinematic action of imbedded particles or asperities rolling and spinning at the ball or roller

contact points. *Note:* The length and curvature of these marks depend on the degree of spinning and on the distance from the spinning axis of the rolling element (Ref 3m).

- **Kingsbury bearing**
  - See tilting-pad bearing .
- **knock**
  - In a spark ignition engine, uneven burning of the air/fuel charge that causes violent, explosive combustion and an audible metallic hammering noise. *Note:* Knock results from premature ignition of the last part of the charge to burn (Ref 8).
- **Knoop (microindentation) hardness number**
  - Developed by F.C. Knoop in the late 1930s, it is the numerical value of microindentation hardness obtained using the Knoop indenter (diamond) and calculated as follows:

$$HK_P = 139.6 \frac{P}{d^2}$$

  - where HK_P (units of GPa) is the ASTM symbol used to represent Knoop microindentation hardness number (with the subscript indicating applied indenter load), *P* is the applied load (in grams), and *d* is the length of the long diagonal of the impression (in micrometers). *Note:* To obtain HK in kg/mm^2, replace 139.6 by 14229.0. See ASTM Standard Test Method E 384 for further details.
- **L**
- **L_{10}-life**
  - See rating life .
- **lacquer**
  - In lubrication, a deposit resulting from the oxidation and/or polymerization of fuels and lubricants when exposed to high temperatures. *Note:* Softer deposits are described as varnishes or gums (Ref 1).
- **lapping**
  - (1) A surface finishing process involving motion against an abrasive embedded in a soft metal (Ref 1). (2) Rubbing two surfaces together, with or without abrasives, for the purpose of obtaining extreme dimensional accuracy or superior surface finish (Ref 11).
- **large-end bearing**
  - See big-end bearing .
- **layer bearing**
  - A bearing constructed in layers (Ref 1). See also bimetal bearing and trimetal bearing .
- **layer-lattice material**
  - Any material having a layerlike crystal structure, but particularly solid lubricants of this type (Ref 1).
- ***L/D* ratio**
  - In bearing technology, the ratio of the axial length of a plain bearing to its diameter (Ref 1m).
- **leaded bronzes**
  - See bearing bronzes .
- **lemon bearing (elliptical bearing)**
  - A two-lobed bearing (Ref 1m).
- **light fraction**
  - The first liquid produced during the distillation of a crude oil (Ref 1).
- **limiting static friction**
  - The resistance to the force tangential to the interface that is just sufficient to initiate relative motion between two bodies under load. *Note:* The term static friction, which properly describes a tangential resistance called into operation by a force less than this, should not be substituted for limiting static friction (Ref 1).
- **linishing**
  - A method of finishing by grinding on a continuous abrasive belt (Ref 1).
- **liquid impact erosion**
  - See erosion (erosive wear) .
- **liquid impingement erosion**

- See erosion (erosive wear) .

- **little-end bearing**
  - A bearing at the smaller (piston) end of a connecting rod in an engine (Ref 1). See also big-end bearing .
- **load**
  - In tribology, the force applied normal to the surface of one body by another contacting body or bodies. *Note:* The term normal force is more precise and therefore preferred; however, the term normal load is also in use. If applied vertically, the load can be expressed in mass units, but it is preferable to use force units such as newtons (N).
- **load-carrying capacity (of a lubricant)**
  - (1) The maximum load that a sliding or rolling system can support without failure (Ref 1m). (2) The maximum load or pressure that can be sustained by a lubricant (when used in a given system under specific conditions) without failure of moving bearings or sliding contact surfaces as evidenced by seizure or welding (Ref 11).
- **loading**
  - (1) The filling of an abrasive paper or other bound abrasive material by abraded material from a second body. (2) The concentration of abrasive bound in a matrix material or added to a lap wheel as a loose abrasive.
- **lobed bearing**
  - A journal bearing with two or more lobes around its periphery produced by machining or by elastic distortion to increase stability or to provide adjustable clearance (Ref 1).
- **low-stress abrasion**
  - A form of abrasion in which relatively low contact pressures on the abrading particles or protuberances cause only fine scratches and microscopic cutting chips to be produced (Ref 3m). See also high-stress abrasion .
- **lubricant**
  - Any substance interposed between two surfaces in relative motion for the purpose of reducing the friction or wear between them (Ref 1). *Note:* The above definition implies intentional addition of a substance to an interface; however, species such as oxides and tarnishes on certain metals can also act as lubricants even though they were not added to the system intentionally.
- **lubricant compatibility**
  - See compatibility (lubricant) .
- **lubrication**
  - The reduction of frictional resistance and wear, or other forms of surface deterioration, between two load-bearing surfaces by the application of a lubricant (Ref 1).
- **lubrication regimes**
  - Ranges of operating conditions for lubricated tribosystems that can be distinguished by their frictional characteristics and/or by the manner and amount of separation of the bearing surfaces. See also boundary lubrication , elastohydrodynamic lubrication , full-film lubrication , hydrodynamic lubrication , and quasi-hydrodynamic lubrication .
- **lubricious (lubricous)**
  - Relating to a substance or surface condition that tends to produce relatively low friction.
- **lubricity**
  - The ability of a lubricant to reduce wear and friction, other than by its purely viscous properties (Ref 1).
- **M**
- **macrohardness test**
  - A term applied to such hardness testing procedures as the Rockwell, Brinell, or macro-Vickers (high-load) hardness tests to distinguish them from microindentation hardness tests such as the Knoop or Vickers tests. See also microindentation hardness number .
- **macroslip**
  - A type of sliding in which all points on one side of the interface are moving relatively to those on the other side in a direction parallel to the interface. *Note:* The term macroslip is sometimes used to denote macroslip velocity. This usage is not recommended (Ref 1). See also microslip .
- **magnetic bearing**
  - A type of bearing in which the force that separates the relatively moving surfaces is produced by a magnetic field.
- **magnetic seal**

- A seal that uses magnetic material, instead of springs or bellows, to provide the closing force (Ref 4).
- **magnetohydrodynamic lubrication**
  - Hydrodynamic lubrication in which a significant force contribution arises from electromagnetic interaction. *Note:* Magnetohydrodynamic bearings have been proposed for very high-temperature operation, for example, in liquid sodium (Ref 1).
- **magnetostrictive cavitation test device**
  - A vibratory cavitation test device driven by a magnetostrictive transducer (Ref 2).
- **main bearing**
  - A bearing supporting the main power-transmitting shaft (Ref 1).
- **mass concentration (in a slurry)**
  - The mass of solid particles per unit mass of mixture, expressed in percent (Ref 13).
- **maximum erosion rate**
  - The maximum instantaneous erosion rate in a test that exhibits such a maximum followed by decreasing erosion rates. *Note:* Occurrence of such a maximum is typical of many cavitation and liquid impingement tests. In some instances it occurs as an instantaneous maximum; in others it occurs as a steady-state maximum that persists for some time (Ref 2).
- **maximum rate period**
  - In cavitation and liquid impingement erosion, a stage following the acceleration period, during which the erosion rate remains constant (or nearly so) at its maximum value (Ref 2).
- **mechanical activation**
  - The acceleration or initiation of a chemical reaction by mechanical exposure of a nascent solid surface. *Note:* Metal cutting is an effective method of exposing large areas of fresh surface.
- **mechanical seal**
  - See face seal .
- **mechanical stability (of a grease)**
  - Grease shear stability tested in a standard rolling tester (Ref 1).
- **mechanical wear**
  - Removal of material due to mechanical processes under conditions of sliding, rolling, or repeated impact. *Note:* The term mechanical wear includes adhesive wear , abrasive wear , and fatigue wear (Ref 1). Compare with corrosive wear and thermal wear .
- **melt lubrication**
  - Lubrication provided by steady melting of a lubricating species. Also known as phase-change lubrication (Ref 1m).
- **metallic wear**
  - Typically, wear due to rubbing or sliding contact between metallic materials that exhibits the characteristics of severe wear, for example, significant plastic deformation, material transfer, and indications that cold welding of asperities possibly has taken place as part of the wear process. See also adhesive wear and severe wear .
- **metallurgical burn**
  - Modification of the microstructure near the contact surface due to frictional temperature rise (Ref 1m).
- **metallurgical compatibility**
  - See compatibility (metallurgical) .
- **microhardness number**
  - A commonly used term for the more technically correct term microindentation hardness number .
- **microindentation**
  - (1) In hardness testing, the small residual impression left in a solid surface when an indenter, typically a pyramidal diamond stylus, is withdrawn after penetrating the surface. *Notes:* Typically, the dimensions of the microindentations are measured to determine microindentation hardness number, but newer methods measure the displacement of the indenter during the indentation process to use in the hardness calculation. The precise size required to qualify as a "microindentation" has not been clearly defined; however, typical measurements of the diagonals of such impressions range from approximately 10 to 200 μm, depending on normal force and material. (2) The process of indenting a solid surface, using a hard stylus of prescribed geometry and under a slowly applied normal force, usually for the purpose of determining its microindentation hardness number. See also Knoop (microindentation) hardness number , microindentation hardness number , and Vickers (microindentation) hardness number .

- **microindentation hardness number**
  - A numerical quantity, usually stated in units of pressure, that expresses the resistance to penetration of a solid surface by a hard indenter of prescribed geometry and under a specified, slowly applied normal force. *Notes:* The prefix "micro" indicates that the indentations produced are typically between 10.0 and 200 μm across. See also Knoop (microindentation) hardness number , nanohardness test , and Vickers (microindentation) hardness number .
- **microslip**
  - Small relative tangential displacement in a contacting area at an interface, when the remainder of the interface in the contacting area is not relatively displaced tangentially. *Notes:* Microslip can occur in both rolling and stationary contacts. (2) The term microslip is sometimes used to denote the microslip velocity. This usage is not recommended (Ref 1). See also macroslip and slip .
- **mild wear**
  - A form of wear characterized by the removal of material in very small fragments. *Notes:* Mild wear is an imprecise term, frequently used in research, and contrasted with severe wear . In fact, the phenomena studied usually involve the transition from mild to severe wear and the factors that influence this transition. Mild wear may be appreciably greater than can be tolerated in practice. With metallic sliders, mild wear debris usually consists of oxide particles (Ref 1m). See also normal wear and severe wear .
- **Miller number**
  - A measure of slurry abrasivity as related to the instantaneous mass-loss rate of a standard metal wear block at a specific time on the cumulative abrasion-corrosion time curve (Ref 13). See also slurry abrasion response number .
- **mineral oil**
  - A refined hydrocarbon oil without animal or vegetable additives (Ref 1m).
- **mist lubrication**
  - Lubrication by an oil mist produced by injecting oil into a gas stream (Ref 1).
- **Mitchell bearing**
  - See tilting-pad bearing .
- **mixed lubrication**
  - See quasi-hydrodynamic lubrication .
- **Mohs hardness**
  - The hardness of a body according to a scale proposed by Mohs, based on ten minerals, each of which would scratch the one below it (Ref 1). These minerals, in decreasing order of hardness, are:

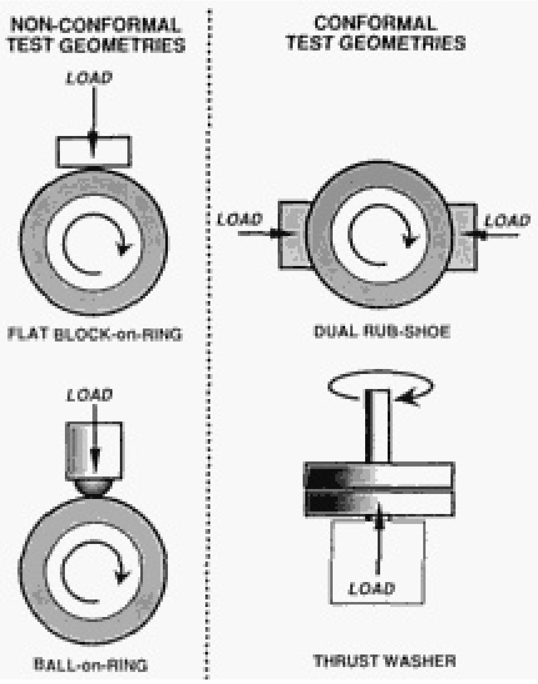
Diamond	10
Corundum	9
Topaz	8
Quartz	7
Orthoclase (feldspar)	6
Apatite	5
Fluorite	4
Calcite	3
Gypsum	2
Talc	1

- **molecular seal**
  - A seal that is basically of the windback type, but that is used for sealing vapors or gases. *Note:* Because of this use, the grooves and lands are dimensioned differently from those of a windback seal (Ref 4).
- **multigrade oil**
  - An oil having relatively little change in viscosity over a specified temperature range (Ref 1).
- **N**

- **nanohardness test**
  - An indentation hardness testing procedure, usually relying on indentation force versus tip displacement data, to make assessments of the resistance of surfaces to penetrations of the order of 10 to 1000 nm deep. *Notes:* The prefix "nano-" normally would imply hardnesses one thousand times smaller than "microhardness"; however, use of this prefix was primarily designed as a means to distinguish this technique from the more traditional microindentation hardness procedures. Most nanohardness testing procedures use three-sided pyramidal diamond indenters first described by Berkovich.
- **nascent surface**
  - A completely uncontaminated surface, produced for example by cleavage fracture under ideal vacuum conditions (Ref 1).
- **Navier-Strokes equations**
  - Basic equations of hydrodynamic lubrication (Ref 1).
- **neat oil**
  - Hydrocarbon oil with or without additives, used undiluted. *Note:* This term is used particularly in metal cutting to distinguish these fluids from soluble oils (emulsions) (Ref 1m).
- **needle bearing**
  - A bearing in which the relatively moving parts are separated by long, thin rollers that have a length-to-diameter ratio exceeding 5.0 (Ref 1m).
- **net positive suction head**
  - The difference between total pressure and vapor pressure in a fluid flow, expressed in terms of equivalent height of fluid, or "head," by the following equation:

$$\text{NPSH} = \left(\frac{P_0}{w}\right) + \left(\frac{V^2}{2g}\right) - \left(\frac{P_v}{w}\right)$$

  - where P_o is the static pressure, P_v is the vapor pressure, V is the flow velocity, w is the specific weight of fluid, and g is the gravitational acceleration. *Note:* This quantity is used in pump design as a measure of the tendency for cavitation to occur at the pump inlet. It can be related to the cavitation number (Ref 2).
- **neutral oil**
  - A lubricating oil obtained by distillation, not treated with acid or with alkali (Ref 1).
- **Newt**
  - The former English unit of kinematic viscosity (Ref 1).
- **Newtonian fluid**
  - A fluid exhibiting Newtonian viscosity wherein the shear stress is proportional to the rate of shear (Ref 1m). Compare with dilatant , rheopectic material , and thixotropy .
- **nip**
  - In a bearing, the amount by which the outer circumference of a pair of bearing shells exceeds the inner circumference of the housing (Ref 1). Also known as crush .
- **NLGI number**
  - Abbreviation for the National Lubricating Grease Institute number, which is the numerical classification of the consistency of greases, based on the ASTM D 217 test (Ref 1).
- **nominal area (of contact)**
  - The area bounded by the periphery of the region in which macroscopic contact between two solid bodies is occurring. *Note:* This is often taken to mean the area enclosed by the boundaries of a wear scar, even though the real area of contact, in which the solids are touching instantaneously, is usually much smaller. See also area of contact and apparent area of contact .
- **nonconformal surfaces**
  - (1) Surfaces whose centers of curvature are on the opposite sides of the interface, as in rolling-element bearings or gear teeth (Ref 1). (2) In wear testing, a geometric configuration in which a "point" or "line" of contact is initially established between specimens before the test is started. *Note:* Examples of nonconformal contacts are ball-on-ring and flat block-on-ring geometries (tangent to the circumferential surface). When wear occurs, the nominal area of contact tends to increase.



Nonconformal. (left) and conformal (right) test geometries

- **noncontact bearing**
  - A bearing in which no solid contact occurs between relatively moving surfaces. *Note:* Strictly speaking, a bearing in which full-film lubrication is occurring would be considered a noncontact bearing; however, this term is more typically applied to gas bearings and magnetic bearings. See also gas lubrication and magnetic bearing .
- **non-Newtonian viscosity**
  - The apparent viscosity of a material in which the shear stress is not proportional to the rate of shear (Ref 1).
- **nonsoap grease**
  - A grease made with a thickener other than soap, such as clay or asbestos (Ref 1).
- **normal force**
  - See load .
- **normal load**
  - See load .
- **normal wear**
  - Loss of material within the design limits expected for the specific intended application. *Note:* The concept of normal wear depends on economic factors, such as the expendability of a worn part.
- **normalized erosion resistance**
  - The volume loss rate of a specified reference material divided by the volume loss rate of a test material obtained under similar testing and analysis conditions. *Notes:* "Similar testing and analysis conditions" means that the volume loss rates of the two materials are determined at the corresponding portions of the erosion rate-time pattern, for example, the maximum erosion rate or the terminal erosion rate. A recommended complete wording has the form: "the normalized erosion resistance of (test material) relative to the (reference material) based on (criterion of data analysis) is (numerical value)" (Ref 2m).
- **O**
- **Ocvirk number**
  - A dimensionless number used to evaluate the performance of journal bearings, and defined by the following equation:

$$\text{Ocvirk number} = \frac{P}{\eta U}\left(\frac{c}{r}\right)^2\left(\frac{d}{b}\right)^2$$

- where *P* is the load per unit width, η is the dynamic viscosity, *U* is the surface velocity, *c* is the radial clearance, *r* is the bearing radius, *b* is the bearing length, and *d* is the bearing diameter. *Note:* This number may be used in its inverted form and is related to the Sommerfeld number (Ref 1m).

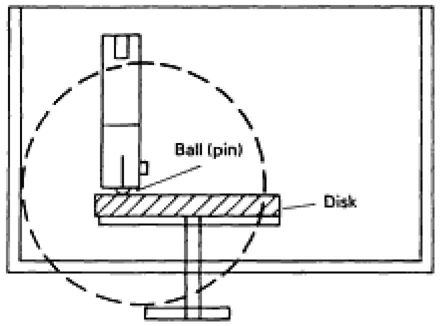
- **oil**
  - A liquid of vegetable, animal, mineral, or synthetic origin that feels slippery to the touch.
- **oil cup**
  - A device connected to a bearing that uses a wick, valve, or other means to provide a regulated flow of lubricant (Ref 1).
- **oil fog lubrication**
  - See mist lubrication .
- **oil groove**
  - A channel or channels cut in a bearing to improve oil flow through the bearing. *Note:* A similar groove may be used for grease-filled bearings (Ref 1).
- **oil pocket**
  - A depression designed to retain oil in a sliding surface (Ref 1).
- **oil ring lubrication**
  - A system of lubrication for horizontal shafts. A ring of larger diameter rotates with the shaft and collects oils from a container beneath (Ref 1).
- **oil starvation**
  - A condition in which a bearing, or other tribocomponent, receives an inadequate supply of lubricant (Ref 1m).
- **oil whirl**
  - Instability of a rotating shaft associated with instability in the lubricant film. *Note:* Oil whirl should be distinguished from shaft whirl, which depends only on the stiffness of the shaft (Ref 1).
- **oiliness**
  - See lubricity .
- **overbasing**
  - A technique for increasing the basicity of lubricants. *Note:* Overbased lubricants are used to assist in neutralizing acidic oxidation products (Ref 1).
- **oxidative wear**
  - A corrosive wear process in which chemical reaction with oxygen or oxidizing environment predominates. *Note:* Usually, oxidative wear is a form of mild wear , but it can sometimes be serious, for example, in liquid oxygen or liquid fluorine (Ref 1).
- **P**
- **pad lubrication**
  - A system of lubrication in which the lubricant is delivered to a bearing surface by a pad of felt or similar material (Ref 1).
- **partial (journal) bearing**
  - A bearing in which the bore extends not more than half the circumference of the journal (Ref 1).
- **partial hydrodynamic lubrication**
  - See quasi-hydrodynamic lubrication .
- **pedestal bearing**
  - A bearing that is supported on a column or pedestal rather than on the main body of the machine (Ref 1).
- **peening wear**
  - Removal of material from a solid surface caused by repeated impacts on very small areas (Ref 1m).
- **penetration (of a grease)**
  - The depth in $\frac{1}{10}$ mm that a standard cone penetrates the sample in a standard cup under prescribed conditions of weight, time (5 s), and temperature (25 °C, or 77 °F). *Notes:* The result depends on whether or not the grease has been subjected to shear. In unworked penetration, the

grease is transferred with as little deformation as possible or is tested in its container. In worked penetration, the grease is subjected to 60 double strokes in a standard device. In prolonged worked penetration, the grease is worked for a specified period before the 60 strokes. The results may be quoted as penetration number or penetration value (Ref 1m).

- **penetration hardness number**
  - Any numerical value expressing the resistance of a body to the penetration of a second, usually harder, body.
- **penetrometer**
  - In grease technology, an instrument for measuring the consistency of a grease by allowing a cone to penetrate into the grease under controlled conditions (Ref 1). See also penetration (of a grease) .
- **petroleum oil**
  - See mineral oil .
- **Petroff equation**
  - An equation describing the viscous power loss in a concentric bearing full of lubricant. The resisting torque on the shaft (T_o = shear stress × shaft radius × bearing area) is given by:

$$T_0 = \frac{\pi^2 \eta NLD^2}{2c}$$

  - where π = 3.1416, η is the dynamic viscosity, N is the shaft speed, L is the bearing length, D is the shaft diameter, and $2c$ is the diametral clearance (Ref 1m).
- **phase-change lubrication**
  - See melt lubrication .
- **phosphor bronzes**
  - See bearing bronzes .
- **pin-on-disk machine**
  - A tribometer in which one or more relatively moving styli (that is, the "pin" specimen) is loaded against a flat disk specimen surface such that the direction of loading is parallel to the axis of rotation of either the disk or the pin-holding shaft, and a circular wear path is described by the pin motion. *Notes:* The typical pin-on-disk arrangement resembles that of a traditional phonograph. Either the disk rotates or the pin specimen holder rotates so as to produce a circular path on the disk surface. An arrangement wherein the pin specimen is loaded against the curved circumferential surface of a flat disk is not generally considered to be a pin-on-disk machine.



Pin-on-disk machine test configuration

- **piston-pin bearing**
  - See little-end bearing .
- **pivot bearing**
  - An axial-load, radial-load type bearing that supports the end of a rotating shaft or pivot (Ref 1).
- **pivoted-pad bearing**
  - See tilting-pad bearing .

- **plain bearing**
  - Any simple sliding-type bearing, as distinct from pad- or rolling-type bearings (Ref 1).
- **plain journal bearing**
  - A plain bearing in which the relatively sliding surfaces are cylindrical and in which there is relative angular motion. *Note:* One surface is usually stationary and the force acts perpendicularly to the axis of rotation (Ref 1).
- **plain thrust bearing**
  - A plain bearing of the axial-load type, with or without grooves (Ref 1).
- **plastohydrodynamic lubrication**
  - A condition of lubrication in which the friction and film thickness between two bodies in relative motion are determined by plastic deformation of the bodies in combination with the viscous properties of the lubricant at the prevailing pressure, temperature, and rate of shear (Ref 1). Compare with elastohydrodynamic lubrication .
- **plowing (ploughing)**
  - The formation of grooves by plastic deformation of the softer of two surfaces in relative motion. *Notes:* The contribution of plowing to frictional resistance is to be distinguished from the contribution of shearing, especially in the classical hemisphere-on-flat model system. In plowing, local solid-phase welding is ignored (Ref 1m).
- **pocket**
  - In a rolling-element bearing, the portion of the case that is shaped to receive the rolling element (Ref 1). Compare with oil pocket .
- **pocket-thrust bearing**
  - An externally pressurized thrust bearing having three or more hydrostatic pads with central relieved chambers of pockets supplied with pressurized oil (Ref 1).
- **poise**
  - (P). The centimeter-gram-second (cgs) unit of dynamic viscosity (Ref 1).
- **Poisseuille**
  - (P1). The meter-kilogram-second (mks) or Système International d'Unités (SI) unit of dynamic viscosity. *Note:* 1 Pl = 10 P (Ref 1).
- **polishing**
  - A surface-finishing process utilizing successive grades of abrasive (Ref 1).
- **polishing wear**
  - An extremely mild form of wear for which the mechanism has not been clearly identified, but that may involve extremely fine-scale abrasion, plastic smearing of micro-asperities, and/or tribochemical material removal.
- **porous bearing**
  - A bearing made from porous material such as compressed and sintered metal powder. *Note:* The pores may act as reservoirs or passages for supplying fluid lubricant, or the bearing may be impregnated with solid lubricant (Ref 1).
- **positive-contact bushing**
  - A bushing, the inside diameter of which has direct contact with the outside diameter of a shaft or sleeve. *Note:* Radial or axial clearances are provided in the housing (Ref 4m).
- **positive-contact seal**
  - A seal, the primary function of which is achieved by one surface mating with another. *Note:* Examples include lip, circumferential, and face-type seals (Ref 4m). See also face seal .
- **pour point**
  - The lowest temperature at which a lubricant can be observed to flow under specified conditions (Ref 1).
- **pour-point depressant**
  - An additive that lowers the pour point of a lubricant (Ref 1m).
- **pressure-viscosity coefficient**
  - The slope of a graph showing variation in the logarithm of viscosity with pressure. *Note:* The use of the term pressure-viscosity coefficient assumes a linear relationship (Barus equation) (Ref 1).
- **pressure lubrication**
  - A system of lubrication in which the lubricant is supplied to the bearing under pressure (Ref 1).
- **pressurized gas lubrication**
  - A system of lubrication in which a gaseous lubricant is supplied under sufficient external pressure to separate the opposing surfaces by a gas film (Ref 1).

- **primary leakage**
  - In seals, the leakage of a mechanical seal, with the fluid escaping from the region between the end faces of the primary sealing elements (Ref 4).
- **protuberances**
  - See asperities .
- **prow formation**
  - See wedge formation .
- **pseudoplastic behavior**
  - A decrease in viscosity with increasing shear stress (Ref 1). Compare with dilatant , rheopectic material , and thixotropy .
- **pumping efficiency**
  - In a bearing, the ratio of actual oil flow to the maximum theoretical flow for a bearing with a 180° oil film operating at an eccentricity ratio of unity (Ref 1).
- **pusher-type seal**
  - A mechanical seal in which the secondary seal is pushed along the shaft or sleeve to compensate for face wear (Ref 4).
- ***PV* factor**
  - The product of bearing pressure and surface velocity traditionally expressed in terms of (lb/in.^{2}) × (ft/min); the ISO equivalent is Pa · m/s (Ref 1m).
- **Q**
- **quasi-hydrodynamic lubrication**
  - A loosely defined regime of lubrication, especially in metalworking, where thin-film lubrication predominates (Ref 1).
- **quench**
  - A neutral fluid that is introduced into a seal chamber or cavity for the purpose of diluting fluid that may have leaked through a seal (Ref 4).
- **quenching oil**
  - Oil used for quenching metals during a heat-treating operation.
- **R**
- **race (or raceway)**
  - The groove or path in which the rolling elements in a rolling-contact bearing operate (Ref 1m).
- **radial lip seal**
  - A radial type of seal that features a flexible sealing member, referred to as a lip. *Note:* The lip is usually of an elastomeric material. It exerts radial sealing pressure on a mating shaft in order to retain fluids and/or exclude foreign matter (Ref 4m).
- **radial-load bearing**
  - A bearing in which the load acts in a radial direction with respect to the axis of rotation (Ref 1).
- **rain erosion**
  - A form of liquid impingement erosion in which the impinging liquid particles are raindrops. *Note:* This form of erosion is of particular concern to designers and material selectors for external surfaces of rotary-wing and fixed-wing aircraft. See also erosion (erosive wear) .
- **rating life**
  - Currently the fatigue life in millions of revolutions or hours at a given operating speed that 90% of a group of substantially identical rolling-element bearings will survive under a given load. *Note:* The 90% rating life is frequently referred to as L_{10}-life or B_{10}-life (Ref 1). From Ref 6, the rating life in revolutions can be obtained from:

$$L_{10} = \left(\frac{C}{P}\right)^K \times 10^6$$

  - where C is the basic load rating (in pounds), P is the equivalent radial load (in pounds), and K is a constant (3 for ball bearings and 10/3 for roller bearings). For a rating life in hours, use:

$$L_{10} = \frac{16700}{N}\left(\frac{C}{P}\right)^K$$

- where *N* is the rotational speed in rev/min.

- **Rayleigh step bearing**
  - A stepped-pad bearing having one step only in each pad (Ref 1).
- **real area of contact**
  - In tribology, the sum of the local areas of contact between two solid surfaces, formed by contacting asperities, that transmit the interfacial force between the two surfaces (Ref 2). Contrast with apparent area of contact .
- **red mud**
  - See cocoa .
- **Redwood viscosity**
  - A commercial measure of viscosity expressed as the time in seconds required for 50 cm^3 of a fluid to flow through a tube of 10 mm length and 1.5 mm diameter at a given temperature. *Note:* It is recommended that fundamental viscosity units should be used (Ref 1).
- **Rehbinder effect**
  - Modification of the mechanical properties at or near the surface of a solid, attributable to interaction with a surfactant (Ref 1).
- **repeated impact**
  - See impingement .
- **restrictor rings**
  - Rings, usually faced with white metal , placed outside a bearing to prevent fluid from being discharged (Ref 1).
- **retainer**
  - See cage .
- **Reyn**
  - The former English unit of dynamic viscosity (Ref 1).
- **Reynold's equation**
  - A basic equation of hydrodynamic lubrication (Ref 1).
- **rheodynamic lubrication**
  - A regime of lubrication in which the rheological (non-Newtonian) properties of the lubricant predominate. *Note:* This term is especially applied to lubrication with grease (Ref 1).
- **rheology**
  - The study of the deformation and/or flow of matter, in terms of stress, strain, temperature, and time (Ref 1).
- **rheopectic material**
  - A material that shows an increase in viscosity with time under a constant shear stress. After removal of the shear stress, the viscosity slowly returns to its original level (Ref 1). Compare with thixotropy .
- **ridging (wear)**
  - A deep form of scratching in parallel ridges usually caused by plastic flow of the subsurface layer (Ref 1).
- **ring seal**
  - A piston ring-type seal that assumes its sealing position under the pressure of the fluid to be sealed (Ref 4).
- **ripple formation (rippling)**
  - Formation of periodic ridges and valleys transverse to the direction of motion (Ref 1).
- **Rockwell hardness number**
  - A number derived from the net increase in the depth of impression as the load on a penetrator is increased from a fixed minor load to a major load and then returned to the minor load (Ref 11). *Note:* Various scales of Rockwell hardness numbers have been developed based on the hardness of the materials to be evaluated. The scales are designated by alphabetic suffixes to the hardness designation, for example, HRB, HRC. A detailed description of the Rockwell hardness testing procedure can be found in ASTM Standard E 40.

- **Rockwell superficial hardness test**
  - The same test as used to determine the Rockwell hardness number except that smaller minor and major loads are used (Ref 11m).
- **roller bearing**
  - A bearing in which the relatively moving parts are separated by rollers (Ref 1).
- **rolling**
  - (pure rolling with no sliding and no spin). A motion of two relatively moving bodies, of opposite curvature, whose surface velocities in the common contact area are identical with regard to both magnitude and direction (Ref 1m). See also sliding and spin .
- **rolling-contact fatigue**
  - Repeated stressing of a solid surface due to rolling contact between it and another solid surface or surfaces. *Note:* Continued rolling-contact fatigue of bearing surfaces may result in rolling-contact damage in the form of subsurface fatigue cracks and/or material pitting and spallation.
- **rolling-contact wear**
  - Wear to a solid surface that results from rolling contact between that surface and another solid surface or surfaces.
- **rolling-element bearing**
  - A bearing in which the relatively moving parts are separated by balls, rollers, or needles (Ref 1).
- **rolling velocity**
  - See sweep velocity .
- **rotary seal**
  - A mechanical seal that rotates with a shaft and is used with a stationary mating ring (Ref 4).
- **rubbing bearing**
  - A bearing in which the relatively moving parts slide without deliberate lubrication (Ref 1).
- **rugosities**
  - See asperities .
- **run-in**
  - (noun). (1) In tribology, an initial transition process occurring in newly established wearing contacts, often accompanied by transients in coefficient of friction, wear rate, or both, that are uncharacteristic of the given tribological system's long-term behavior (Ref 2). (2) In seals, the period of initial operation during which the seal-lip wear rate is greatest and the contact surface is developed (Ref 4m).
- **run in**
  - (verb). In tribology, to apply a specified set of initial operating conditions to a tribological system in order to improve its long-term frictional or wear behavior, or both (Ref 2). *Note:* The run-in may involve conditions either more severe or less severe than the normal operating conditions of the tribosystem, and may also involve the use of special lubricants and/or surface chemical treatments. See also break-in .
- **running-in**
  - The process by which machine parts improve in conformity, surface topography, and frictional compatibility during the initial stage of use. *Note:* Chemical processes, including formation of an oxide skin, and metallurgical processes, such as strain hardening, may contribute (Ref 1).
- **S**
- **saponify**
  - (1) To convert into soap. (2) To subject to, or to undergo, saponification (Ref 5).
- **saponification number**
  - A measure of the amount of constituents of petroleum that will easily saponify under test conditions, determined by the number of milligrams of sodium hydroxide that is consumed by 1 g of oil under test conditions. *Note:* Saponification number is a measure of fatty materials compounded in an oil (Ref 5m).
- **sauter mean diameter**
  - The diameter of a drop that has the same ratio of volume-to-surface area as the ratio of total volume-to-total surface area in a distribution of drops, as computed from the equation:

$$\text{SMD} = \Sigma_i n_i d_i^3 / \Sigma_i n_i d_i^2$$

- where *i* is a sampling size interval, d_i is the drop diameter, and n_i is the number of drops in that interval (Ref 2).
- **Saybolt Universal Viscosity**
  - A commercial measure of viscosity expressed as the time in seconds required for 60 cm^3 of a fluid to flow through the orifice of the Standard Saybolt Universal Viscometer at a given temperature under specified conditions (Ref 1).
- **scabbing**
  - (1) In wear, a loosely used term referring to the formation of bulges in the surface (Ref 1). (2) In fracture mechanics, it is identical with spalling (Ref 1).
- **scaling**
  - The formation of a thick oxide on the surface of a heated metal. *Note:* Scaling is to be distinguished from rusting, which involves the formation of hydrated oxides (Ref 1).
- **scoring**
  - (1) The formation of severe scratches in the direction of sliding. *Notes:* Scoring may be due to local solid-phase welding or to abrasion. In the United States, the term scuffing is sometimes used as a synonym for scoring. Minor damage should be called scratching rather than scoring (Ref 1). (2) In tribology, a severe form of wear characterized by the formation of extensive grooves and scratches in the direction of sliding (Ref 2). (3) The act of producing a scratch or narrow groove in a surface by causing a sharp instrument to move along that surface. See also galling .
- **scouring abrasion**
  - Same as abrasion .
- **scraper**
  - An exclusion seal that has metallic or other firm lips or scraping elements. It serves to remove foreign material from a reciprocating shaft (Ref 4).
- **scratch**
  - A groove produced in a solid surface by the cutting and/or plowing action of a sharp particle or protuberance moving along that surface.
- **scratch hardness test**
  - A form of hardness test in which a sharp-pointed stylus or corner of a mineral specimen is traversed along a surface so as to determine the resistance of that surface to cutting or abrasion. *Notes:* The Mohs hardness test is among the most widely used forms of scratch hardness tests, but is mainly applied to mineralogical specimens or abrasives. Other scratch hardness tests involve using diamond cones, pyramids, and spherical tips, but such scratch hardness tests have not been established and standardized to the extent that macro- and microindentation hardness tests have been.
- **scratching**
  - The formation of fine scratches in the direction of sliding. *Notes:* Scratching may be due to asperities on the harder slider or to hard particles between the surface or embedded in one of them. Scratching is to be considered less damaging than scoring (Ref 1). See also abrasion , plowing (ploughing) , and ridging wear .
- **scratching abrasion**
  - See low-stress abrasion .
- **scuffing**
  - (1) Localized damage caused by the occurrence of solid-phase welding between sliding surfaces, without local surface melting. *Notes:* In the United Kingdom, scuffing implies local solid-phase welding only. In the United States, scuffing may include abrasive effects and the term scoring is sometimes used as a synonym (Ref 1). (2) A mild degree of galling that results from the welding of asperities due to frictional heat. The welded asperities break, causing surface degradation (Ref 4). *Note:* In general, the term scuffing has been used in so many different ways that its use should be avoided whenever possible, and instead replaced with a more precise description of the specific type of surface damage being considered.
- **seal**
  - (1) In tribology, a device designed to prevent leakage between relatively moving parts (Ref 1). (2) A device designed to prevent the movement of fluid from one chamber to another, or to exclude contaminants (Ref 4).
- **sealing face**

- The lapped surface of a seal that comes in close proximity to the face of the mating ring of a face seal, thus forming the primary seal. With reference to lip seals, the preferred term is seal contact surface (Ref 4).

- **seal nose**
  - The part of the primary seal ring of a face seal that comes in closest proximity to the mating surface and that, together with the mating surface, forms the primary seal (Ref 4).
- **secondary seal**
  - A device, such as bellows, piston ring, or O-ring, that allows axial movement of the primary seal of a mechanical face seal without undue leakage (Ref 4).
- **seizure**
  - The stopping of relative motion as the result of interfacial friction. *Notes:* Seizure may be accompanied by gross surface welding. The term is sometimes used to denote scuffing (Ref 1m).
- **selective transfer**
  - A process involving the transfer and attachment of a specific species from one surface to the mating surface during sliding. *Notes:* This term is to be distinguished from the general tribological term transfer , which involves the same general process, but which, when used by itself, does not discriminate as to which of the species present in a multiconstituent surface is transferred during sliding. This term is more commonly used in the Russian literature of tribology.
- **self-acting bearing**
  - See gas bearing and self-lubricating bearing .
- **self-aligning bearing**
  - A rolling-element bearing with one spherical raceway that automatically provides compensation for shaft or housing deflection or misalignment (Ref 1). Compare with aligning bearing .
- **self-lubricating bearing**
  - A bearing independent of external lubrication. *Note:* These bearings may be sealed for life after packing with grease or may contain self-lubricating material (Ref 1).
- **self-lubricating material**
  - Any solid material that shows low friction without application of a lubricant. *Notes:* Examples are graphite, molybdenum disulfide, and polytetrafluoroethylene. Taken in a broader context, the term can also refer to a composite material into which a lubricious species has been incorporated (Ref 1m). See also solid lubricant .
- **separator**
  - In rolling-element bearings, the part of a cage that lies between the rolling elements. *Note:* This term is sometimes used as a synonym for cage (Ref 1m).
- **severe wear**
  - A form of wear characterized by removal of material in relatively large fragments. *Notes:* Severe wear is an imprecise term, frequently used in research, and contrasted with mild wear . In fact, the phenomena studied usually involve the transition from mild to severe wear and the factors that influence that transition. With metals, the fragments are usually predominantly metallic rather than oxidic. Severe wear is frequently associated with heavy loads and/or adhesive contact (Ref 1).
- **shaft run-out**
  - Twice the distance that the center of a shaft is displaced from the axis of rotation; that is, twice the eccentricity (Ref 4).
- **shake-down (of surface layers)**
  - The establishment of a state of stress in which no further plastic flow occurs (Ref 1m).
- **shear stability**
  - The ability of a lubricant to withstand shearing without degradation (Ref 1). See also penetration (of a grease) .
- **shear thickening**
  - An increase in viscosity with an increase in shear stress or time (Ref 1). See also shear thinning .
- **shear thinning**
  - A decrease in viscosity with an increase in shear stress or time. *Note:* The decrease in viscosity may be temporary or permanent. The latter happens when the shear stress is sufficiently large to rupture a chemical bond, so that the sheared liquid has a lower viscosity than it had prior to shearing (Ref 1). See also shear thickening .
- **shelling**

- A term used in railway engineering to describe an advanced phase of spalling (Ref 1).
- **skidding**
  - A form of nonuniform relative motion between solid surfaces due to rapid periodic changes in the traction between those surfaces.
- **sleeve bearing**
  - A cylindrical plain bearing used to provide radial location for a shaft, which moves axially. *Note:* Sleeve bearing is sometimes used to denote journal bearing (Ref 1). See also sliding bearing .
- **slide-roll ratio**
  - See slide-sweep ratio .
- **slide-sweep ratio**
  - The ratio of sliding velocity to sweep velocity, for example, in a pair of gears. *Note:* In rolling, the slide-sweep ratio is called the slide-roll ratio (Ref 1).
- **sliding**
  - (pure sliding with no rolling or spin). A motion of two relatively moving bodies, in which their surface velocities in the common contact area are different with regard to magnitude and/or direction (Ref 1). See also rolling , spin , and specific sliding .
- **sliding bearing**
  - A bearing in which predominantly sliding contact occurs between relatively moving surfaces. *Note:* Sliding bearings may be either unlubricated, liquid lubricated, grease lubricated, or solid lubricated.
- **sliding velocity**
  - The difference between the velocities of each of the two surfaces relative to the point of contact (Ref 1).
- **sludge**
  - A coagulated mass, often containing foreign matter, formed at low temperature in combustion engines from oil oxidation residues, carbon, and water (Ref 1m).
- **slumpability**
  - The flow of gravity of a grease in a container, allowing it to feed out into a pump or can. *Note:* Slumpability also influences the leakage of grease from a bearing (Ref 1).
- **slurry**
  - A mixture of solid particles in a liquid, of such consistency so as to be capable of being pumped like a liquid (Ref 13).
- **slurry abrasion response (SAR) number**
  - A measure of the relative abrasion response of any material in any slurry, as related to the instantaneous rate of mass loss of a specimen at a specific time on the cumulative abrasion-corrosion time curve, converted to volume or thickness loss rate (Ref 13). *Notes:* The SAR number is determined by a procedure described in ASTM Standard Test Method G 75. The Miller number , also described in ASTM G 75, is a measure of the abrasivity of a slurry.
- **slurry abrasivity**
  - The relative tendency of a particular moving slurry to produce abrasive and corrosive wear compared with other slurries (Ref 13).
- **slurry erosion**
  - Erosion produced by the movement of a slurry past a solid surface.
- **slushing oil**
  - A mineral oil containing additives that enable it to protect the parts of a machine against rusting (Ref 1).
- **smearing**
  - Mechanical removal of material from a surface, usually involving plastic shear deformation, and redeposition of the material as a thin layer on one or both surfaces (Ref 1m). See also transfer .
- **soap**
  - In lubrication, a compound formed by the reaction of a fatty acid with a metal or metal compound. *Note:* Metallic soaps formed by reaction *in situ* are an important group of boundary lubricants (Ref 1).
- **solid-film lubrication**
  - Lubrication by application of solid lubricants (Ref 1).
- **solid lubricant**
  - Any solid used as a powder or thin film on a surface to provide protection from damage during relative movement, and to reduce friction and wear. *Notes:* Many solid lubricants have a layer

(lamellar) structure, for example, graphite and molybdenum disulfide. Certain solid materials lubricate only in the molten state at the interface. Examples are glass and ice, which are not solid lubricants (Ref 1).

- **soluble oil**
  - A mineral oil containing additives that enable it to form a stable emulsion with water. *Note:* Soluble oils are not soluble in water (Ref 1).
- **Sommerfeld number**
  - A dimensionless number that is used to evaluate the performance of journal bearings. It is numerically defined as follows:

$$\frac{P}{\eta U}\left(\frac{c}{r}\right)^2$$

  - where *P* is the load per unit width, η is the dynamic viscosity, *U* is the surface velocity, *c* is the radial clearance, and *r* is the bearing radius. *Notes:* At lower concentricities it is convenient to use the Sommerfeld number in the form given. Because it tends to infinity as the eccentricity approaches unity, the reciprocal form is frequently used in the case of heavily loaded bearings. The expression:

$$\frac{\eta N}{p}\left(\frac{r}{c}\right)^2$$

  - in which *N* is the frequency of rotation and *p* is the pressure, is sometimes referred to as the Sommerfeld number, particularly in the United States (Ref 1m). See also Hershey number and Ocvirk number .
- **spalling**
  - (1) Separation of particles from a surface in the form of flakes. *Notes:* The term spalling is commonly associated with rolling-element bearings and with gear teeth. Spalling is usually a result of subsurface fatigue and is more extensive than pitting (Ref 1). (2) In tribology, the separation of macroscopic particles from a surface in the form of flakes or chips, usually associated with rolling-element bearings and gear teeth, but also resulting from impact events (Ref 2). See also fatigue wear , scabbing , and shelling .
- **spark erosion**
  - See electrical pitting .
- **specific sliding**
  - The ratio of the algebraic difference between the surface velocities of two bodies in relative motion to their sum (Ref 1).
- **specific wear rate**
  - In journal bearings, the proportionality constant *K* in the equation:

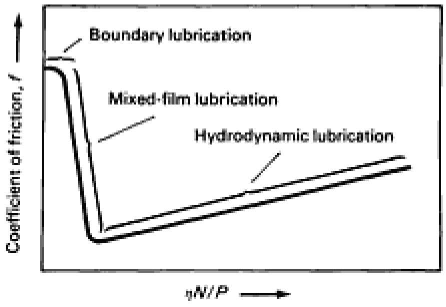
$$h = Kpvt$$

  - where *h* is the radial wear in the bearing, *p* is the apparent contact pressure, *v* is the velocity of the journal, and *t* is the sliding time (Ref 14). *Note:* The constant *K* has also been called the wear factor , but there are other definitions for the term wear factor that do not necessarily refer to journal bearings or derive their meanings from the above equation.
- **spherical bearing**
  - A bearing that is self-aligning by virtue of its partially spherical form (Ref 1).
- **spherical roller bearing**
  - (1) A spherical bearing containing rollers (Ref 1). (2) A roller bearing containing barrel-shaped or hour glass-shaped rollers riding on spherical (concave or convex) races to provide self-aligning capability (Ref 8).

- **spin**
  - In bearings, rotation of a rolling element about an axis normal to the contact surfaces (Ref 1). See also rolling and sliding .
- **spindle oil**
  - An oil of low viscosity used to lubricate high-speed light spindles (Ref 1).
- **split seal**
  - A seal that has its primary sealing elements split in a plane parallel to the axis of the shaft such that, instead of the rings being continuous, they are essentially two semicircles. Modified designs of lip seals feature units with a single lip separation and with one or more separations of the metallic stiffening mechanisms (Ref 4).
- **splash lubrication**
  - A system of lubrication in which the lubricant is splashed onto the moving parts (Ref 1).
- **sponge effect**
  - See squeeze effect .
- **spragging**
  - Intermittent motion arising from design features that allow an increase in tangential force or displacement to produce an increase in normal force (Ref 1).
- **spreader**
  - An axial groove in a plain bearing designed to spread oil along the bearing (Ref 1). See also gutterway .
- **spreader pockets**
  - Depressions in a sliding surface designed to distribute lubricant (Ref 1).
- **squeeze effect**
  - The production of lubricant from a porous retainer by application of pressure. Also known as sponge effect (Ref 1). (2) The persistence of a film of fluid between two surfaces that approach each other in the direction of their common normal (Ref 1).
- **starting torque**
  - The torque that is required for initiating rotary motion (Ref 4).
- **static coefficient of friction**
  - The coefficient of friction corresponding to the maximum friction force that must be overcome to initiate macroscopic motion between two bodies (Ref 2).
- **static equivalent load**
  - (P_o). In rolling-element bearings, the static load which, if applied, would give the same life as that which the bearing will attain under actual conditions of load and rotation (Ref 6m).
- **static friction**
  - See limiting static friction .
- **static load rating**
  - (C_o). In rolling-element bearings, the static load that corresponds to a permanent deformation of rolling element and race at the most heavily stressed contact of 0.00001 of the rolling-element diameter (Ref 6m).
- **static viscosity**
  - See viscosity .
- **stave bearing**
  - A sleeve bearing consisting of several axially held slats or staves on the outer surface of which the bearing material is bonded (Ref 1).
- **step bearing**
  - A plain surface bearing that supports the lower end of a vertical shaft. Other types of bearings may be thus described when they are mounted on a step or bracket (Ref 1). See also Rayleigh step bearing and stepped bearing .
- **stepped bearing**
  - A thrust bearing in which the working face consists of one or more shallow steps. *Notes:* Bearings of this type were originally described by Rayleigh. A distinction should be drawn between a stepped bearing and a step bearing (Ref 1).
- **stern-tube bearing**
  - The final bearing through which a propeller shaft passes in a boat or ship (Ref 1m).
- **stick-slip**
  - A relaxation oscillation usually associated with a decrease in the coefficient of friction as the relative velocity increases. *Notes:* Stick-slip was originally associated with formation and

destruction of interfacial junctions on a microscopic scale. This is often the basic cause. The period depends on the velocity and on the elastic characteristics of the system. Stick-slip will not occur if the static friction is equal to or less than the dynamic friction. The motion resulting from stick-slip is sometimes referred to as jerky motion (Ref 1). See also spragging .

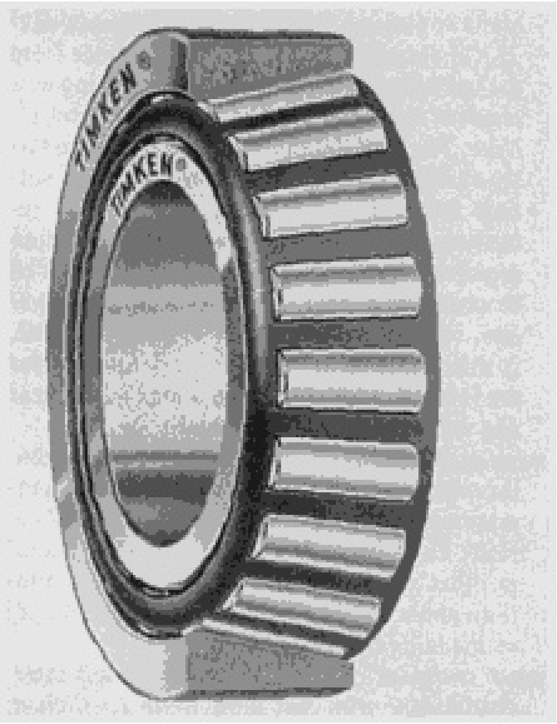
- **stiction**
  - A term sometimes used to signify the condition in which the frictional resistance is sufficient to prevent macroscopic sliding (Ref 1).
- **Stoke (centistoke)**
  - The centimeter-gram-second (cgs) unit of kinematic viscosity (Ref 1).
- **storage stability**
  - A measure of the ability of a lubricant to undergo prolonged periods of storage without showing any adverse conditions due to oxidation, oil separation, contamination, or any type of deterioration (Ref 5).
- **Stribeck curve**
  - A graph showing the relationship between coefficient of friction and the dimensionless number ($\eta N/P$), where η is the dynamic viscosity, *N* is the speed (revolutions per minute for a journal), and *P* is the load per unit of projected area. *Note:* The symbols Z and *v* (linear velocity) may be substituted for η and *N,* respectively (Ref 1).



Stribeck curve: coefficient of friction (and type of lubrication) versus dimensionless variable $\eta N/P$

- **sulfochlorinated lubricant**
  - A lubricant containing chlorine and sulfur compounds, which react with a rubbing surface at elevated temperatures to form a protective film. *Note:* There may be a synergistic effect, producing faster reaction than with sulfur or chlorine additives alone (Ref 1).
- **sulfurized lubricant**
  - A lubricant containing sulfur or a sulfur compound that reacts with a rubbing surface at elevated temperatures to form a protective film. *Note:* The shear strength of the sulfide film formed on ferrous materials is lower than that of the metal but greater than that of the film formed by reaction with a chlorinated lubricant (Ref 1).
- **superficial hardness test**
  - See Rockwell superficial hardness test .
- **surface damage**
  - In tribology, damage to a solid surface resulting from mechanical contact with another substance, surface, or surfaces moving relatively to it and involving the displacement or removal of material. *Note:* In certain contexts, wear is a form of surface damage in which material is progressively removed. In another context, surface damage involves a deterioration of function of a solid surface even though there is no material loss from that surface. Surface damage may therefore precede wear.
- **surface distress**
  - In bearings and gears, damage to the contacting surfaces that occurs through intermittent solid contact involving some degree of sliding and/or surface fatigue. *Note:* Surface distress can occur in numerous forms depending on the conditions under which the bearing or gear was operated and on the nature of the interaction between the contacting surfaces.

- **surfactant**
  - A chemical substance characterized by a strong tendency to form adsorbed interfacial films when in solution, emulsion, or suspension, thus producing effects such as low surface tension, penetration, boundary lubrication, wetting, and dispersing (Ref 1). See also Rehbinder effect .
- **sweating**
  - Exudation of bearing material or lubricant due to high temperature (Ref 1).
- **sweep velocity**
  - The mean of the surface velocities of two bodies at the area of contact. *Notes:* Occasionally the sum of the velocities is quoted instead of the mean. In rolling, the sweep velocity is also called the rolling velocity (Ref 1).
- **syneresis (of a grease)**
  - See bleeding .
- **synthetic oil**
  - Oil produced from chemical synthesis rather than from petroleum. *Note:* Examples are esters, ethers, silicons, silanes, and halogenated hydrocarbons.
- **T**
- **tandem seal**
  - A multiple-seal arrangement consisting of two seals mounted one after the other, with the faces of the seal heads oriented in the same direction (Ref 4).
- **taper section**
  - A section cut obliquely through a surface and prepared metallographically. *Note:* The angle is often chosen to increase the vertical magnification of surface features by a factor of 5 or 10 (Ref 1).
- **tapered land bearing**
  - A thrust bearing containing pads of fixed taper (Ref 1).
- **tapered roller bearing**
  - A rolling-element bearing containing tapered rollers (Ref 1).



Tapered roller bearing (cutaway view). Courtesy of D. Huffman, The Timken Company

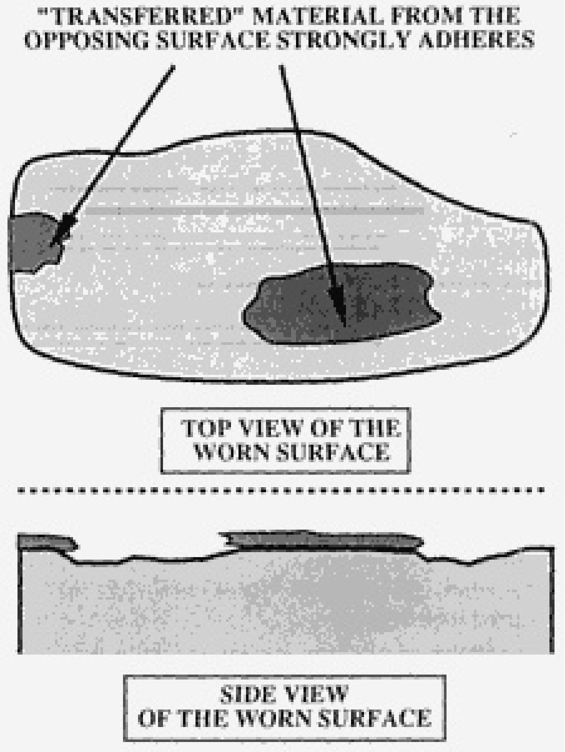
- **Taylor vortices**
  - In a journal bearing, vortices formed in a liquid occupying the annular space between two concentric cylinders (Ref 1).
- **terminal erosion rate**

- The final steady-state erosion rate that is reached (or appears to be approached asymptotically) after the erosion rate has declined from its maximum value. This occurs in some, but not all, cavitation and liquid impingement tests (Ref 2).

- **terminal period**
  - In cavitation and liquid impingement erosion, a stage following the deceleration period, during which the erosion rate has leveled off and remains approximately constant (sometimes with superimposed fluctuations) at a value substantially lower than the maximum rate attained earlier (Ref 2).
- **textile oil**
  - (1) An oil used to lubricate thread or yarn to prevent breakage during spinning and weaving. (2) An oil acceptable for direct contact with fibers during textile production (Ref 1).
- **thermal taper**
  - See thermal wedge .
- **thermal wear**
  - Removal of material due to softening, melting, or evaporation during sliding or rolling. *Notes:* Thermal shock and high-temperature erosion may be included in the general description thermal wear. Wear by diffusion of separate atoms from one body to the other, at high temperatures, is sometimes denoted as thermal wear (Ref 1).
- **thermal wedge**
  - The increase in pressure due to the expansion of the lubricant, for example, in a parallel thrust bearing. *Note:* Thermal distortion of the bearing surfaces may also form a wedge shape. This is referred to as thermal taper (Ref 1).
- **thermoelastic instability**
  - (TEI) In sliding contact, sharp variations in local surface temperatures with the passing of asperities leading to stationary or slowly moving hot spots of significant magnitude (Ref 8).
- **thickener**
  - A solid material dispersed in a liquid lubricant to produce a grease. *Note:* Silica, clays, and metallic soaps are widely used as thickeners (Ref 1).
- **thick-film lubrication**
  - A condition of lubrication in which the film thickness of the lubricant is appreciably greater than that required to cover the surface asperities when subjected to the operating load, so that the effect of the surface asperities is not noticeable (Ref 1). See also thin-film lubrication .
- **thin-film lubrication**
  - A condition of lubrication in which the film thickness of the lubricant is such that the friction and wear between the surfaces is determined by the properties of the surfaces as well as the viscosity of the lubricant. *Notes:* Under thin-film conditions, the coefficient of friction is often 10 to 100 times greater than under thick-film conditions and wear is no longer negligible. Boundary lubrication and elastohydrodynamic lubrication are better-defined terms that should be used where possible (Ref 1).
- **thixotropy**
  - The property of recovering consistency after a decrease as a result of shearing. *Note:* Thixotropic age hardening is a more prolonged process than the instantaneous increase in apparent viscosity as shear rate decreases, which is found in non-Newtonian fluids (Ref 1). See also rheopectic material .
- **thrust bearing**
  - A bearing in which the load acts in the direction of the axis of rotation (Ref 1).
- **tilting-pad bearing**
  - A pad bearing in which the pads are free to take up a position at an angle to the opposing surface according to the hydrodynamic pressure distribution over its surface (Ref 1).
- **torque**
  - As applied to sealing, a resistance to shaft rotation caused by the frictional drag of the seal. It is normally expressed in foot-pound or inch-pound units (Ref 4). *Note:* The metric equivalent is N · m (force times distance).
- **track**
  - The mark made by a seal on the surface with which it mates (Ref 4).
- **tracking pattern**
  - The path a seal ring makes when in rubbing contact with the mating ring or seal plate (Ref 4).
- **traction**

- In rolling contacts, the tangential stress transmitted across the interface. *Notes:* The traction will in general vary from point to point over the contact area. More generally, traction may denote the force per unit area of contact (Ref 1).

- **tractive force**
  - The integral of the tangential surface stress over the area of contact (Ref 1).
- **transfer**
  - In tribology, the process by which material from one sliding surface becomes attached to another surface, possibly as the result of interfacial adhesion. *Notes:* Transfer is usually associated with adhesion, but the possibility of mechanical interlocking adherence, without adhesive bonding, exists in certain occurrences. Material may also back transfer to the surface from which it came. See also selective transfer .



Transfer (schematic)

- **transmission oil**
  - (1) Oil used for transmission of hydraulic power. (2) Oil used to lubricate automobile transmission systems (Ref 1).
- **transition diagram**
  - In tribology, a plot of two or more experimental or operating variables that indicates the boundaries between various regimes of wear or surface damage. *Notes:* The IRG transition diagram is a plot of normal force (ordinate) versus sliding velocity (abscissa), and is used to identify three regions with differing lubrication effectiveness. Various plots have been called transition diagrams, and the context of usage must be established.
- **tribo-**
  - A prefix indicating a relationship to interacting surfaces in relative motion.
- **tribochemistry**
  - The part of chemistry dealing with interacting surfaces in relative motion (Ref 1). *Note:* Tribochemistry broadly encompasses such areas as lubricant chemistry, changes in reactivity of surfaces due to mechanical contact, oxidative wear, and other phenomena.
- **triboelement**
  - A solid body that is bounded by one or more tribosurfaces and that resides within a tribosystem . *Note:* This term is often associated with H. Czichos, whose systems approach to tribology

identified triboelements. For example, in a pin-on-disk tribosystem, the pin is one triboelement and the disk is another.

- **tribology**
  - The science and technology of interacting surfaces in relative motion and of the practices related thereto. *Notes:* This term was introduced in a report by a group set up by the British Department of Education and Science. It was defined in their report "Lubrication (Tribology)" in 1966. The study of wear by erosion or cavitation erosion is included in tribology (Ref 1). (2) Friction, lubrication, and wear science and technology.
- **tribometer**
  - (1) An instrument or testing rig to measure normal and frictional forces of relatively moving surfaces (Ref 1). (2) Any device constructed for or capable of measuring the friction, lubrication, and wear behavior of materials or components.
- **tribophysics**
  - That part of physics dealing with interacting surfaces in relative motion (Ref 1).
- **triboscience**
  - The scientific discipline devoted to the systematic study of interacting surfaces in relative motion. *Note:* Triboscience includes the scientific aspects of tribochemistry , tribophysics , contact mechanics, and materials and surface sciences as related to tribology .
- **tribosurface**
  - Any solid surface whose intermittent, repeated, or continuous contact with another surface or surfaces, in relative motion, results in friction, wear, and/or surface damage. *Note:* The surface of a body subjected to a catastrophic collision would not generally be considered a tribosurface because significant damage to the entire body is involved.
- **tribosystem**
  - Any functional combination of triboelements, including thermal and chemical surroundings.
- **tribotechnology**
  - The aspect of tribology that involves engineering application of triboscience and the design, development, analysis, and repair of components for tribological applications.
- **trimetal bearing**
  - A bearing consisting of three layers. *Note:* Trimetal bearings are usually made of bronze with a white metal facing and a steel backing (Ref 1).
- **trunnion bearing**
  - A bearing used as a pivot to swivel or turn an assembly (Ref 1).
- **turbine oil**
  - An oil used to lubricate bearings in a steam or gas turbine (Ref 1).
- **U**
- **unctuous**
  - A general term expressing the slippery feel of a lubricant when rubbed with the fingers. *Note:* It is recommended that this term not be used in English (Ref 1).
- **uniformly distributed impact test**
  - See distributed impact test .
- **unlubricated sliding**
  - Sliding without lubricant but not necessarily under completely dry conditions (Ref 1). *Note:* Unlubricated sliding is often used to mean "not intentionally lubricated," but surface species such as naturally formed surface oxides and other interfacial contaminants may act in a lubricious manner in nominally unlubricated sliding.
- **V**
- **V-ring seal**
  - A seal consisting of a ring or nested rings that have a V-shaped cross section and that are commonly made from elastomeric material. *Note:* Spring loading is sometimes used to maintain contact between the seal and its mating surface. It is normally used to seal against axial motion (Ref 4m).
- **vacuum residue**
  - The residue from vacuum distillation of crude oil (Ref 1).
- **vapor-phase lubrication**
  - A type of lubrication in which one or more gaseous reactants are supplied to the vicinity of the surface to be lubricated and which subsequently react to form a lubricious deposit on that surface.

Note: The surface to be lubricated may exhibit a catalytic effect to permit formation of the lubricating film.

- **varnish**
  - In lubrication, a deposit resulting from the oxidation and/or polymerization of fuels, lubricating oils, or organic constituents of bearing materials. *Note:* Harder deposits are described as lacquers, softer deposits are described as gums (Ref 1).
- **VI improver**
  - An additive, usually a polymer, that reduces the variation of viscosity with temperature, thereby increasing the viscosity index of an oil (Ref 1).
- **vibratory cavitation**
  - Cavitation caused by the pressure fluctuations within a liquid, induced by the vibration of a solid surface immersed in the liquid (Ref 2).
- **Vickers (microindentation) hardness number**
  - The numerical value of microindentation hardness obtained using the Vickers (diamond) indenter and calculated as follows:

$$HV_P = 18.192\left[\frac{P}{0.5\,(d_1 + d_2)}\right]$$

  - where HV_P (in units of GPa) is the ASTM symbol indicating Vickers microindentation hardness, *P* is the applied load (in grams), and d_1 and d_2 are the two diagonal lengths of the impression (in μm). *Notes:* To obtain the HV_P in kg/mm^2, as was customary in the past, the factor 18.192 is replaced with 1854.4. The Vickers microindentation hardness test is described in ASTM E 384. A higher-load, macro-Vickers test has also been developed.
- **viscoelasticity**
  - A combination of viscous and elastic properties that makes the deformation dependent upon both temperature and strain rate (Ref 1).
- **viscosity**
  - The bulk property of a fluid, semifluid, or semisolid substance that causes it to resist flow. *Notes:* Viscosity is defined by the equation:

$$\eta = \frac{\tau}{(dv/ds)}$$

  - where τ is the shear stress, *v* is the velocity, and *s* is the thickness of an element measured perpendicular to the direction of flow; (*dv*/*ds*) is known as the rate of shear. Newtonian viscosity is often called dynamic viscosity, or absolute viscosity. Kinematic viscosity, or static viscosity (v) is the ratio of dynamic viscosity (η) to density (ρ) at a specified temperature and pressure ($v = \eta/\rho$). It is recommended that the term static viscosity should not be used (Ref 1m).
- **viscosity, units of**
  - For absolute viscosity, the centimeter-gram-second (cgs) unit is the poise (dyne · s/cm^2), the meter-kilogram-second (mks) unit is the Poiseuille (newton · s/m^2), which is equal to the Système International d'Unités (SI) unit (1.0 N · sm^{-1} = 10 poise), and the English unit is the Reyn (lb · s/in.^{2}), which is not commonly used. For kinematic viscosity, the cgs unit is the Stoke (cm^2/s), the SI unit is m^2 · s^{-1} (1.0 m^2 · s^{-1} = 10^6 centistokes), and the English unit is the Newt (in.^{2}/s), which is not commonly used (Ref 1m).
- **viscosity index (VI)**
  - A commonly used measure of the change in viscosity of a fluid with temperature. *Notes:* The higher the viscosity index, the smaller the relative change in viscosity with temperature. Two different indices are used: The earlier usage, according to Dean and Davis, applies to oils having a VI from 0 to 100. Extended VI applies to oils having a VI of at least 100. It compares the oil with a reference oil of VI 100 (Ref 1m).
- **viscous**
  - Possessing viscosity. *Note:* This term is frequently used to imply high viscosity (Ref 1).

- **viscous friction**
  - See fluid friction .
- **W**
- **wear**
  - (1) Damage to a solid surface, generally involving progressive loss of material, that is due to relative motion between that surface and a contacting substance or substances (Ref 2). (2) The progressive loss of substance from the operating surface of a body occurring as a result of relative motion at the surface (Ref 1). *Note:* The first definition would not necessarily exclude scuffing or galling , in which surface material may be displaced but not removed; however, the second definition requires substance to be removed in order to be considered wear. Sometimes surface damage , which may not involve material removal, is differentiated from wear as defined under the first definition. This Handbook treats wear and surface damage individually.
- **wear coefficient**
  - In sliding wear, a dimensionless number that usually represents the proportionality factor *k* in the Archard wear law . *Note:* The term wear coefficient has been used in other ways than that given in this definition, and when encountered, it is best to verify its meaning from context.
- **wear constant**
  - A term sometimes used interchangeably with the term wear coefficient ; however, it should not be used because wear rate is often not constant under different sliding conditions, and, therefore, the term is misleading.
- **wear debris**
  - Particles that become detached in a wear process (Ref 1m).
- **wear factor**
  - A term that has no unique meaning. When encountered, the meaning of this term must be taken from its immediate context. Sometimes, the term is used to refer to the sliding wear rate in terms of $mm^3 N \cdot m$. See also specific wear rate .
- **wear rate**
  - There is no single, standard way to express wear rate. The units used depend on the type of wear and the nature of the tribosystem in which wear occurs. Wear rate can be expressed, for example, as (1) volume of material removed per unit time, per unit sliding distance, per revolution of a component or per oscillation of a body (that is, in sliding wear), (2) volume loss per unit normal force per unit sliding distance (mm^3/N m, which is sometimes called the wear factor), (3) mass loss per unit time, (4) change in a certain dimension per unit time, and (5) relative change in dimension or volume with respect to the same changes in another (reference) substance. The manner of expressing wear rate is sometimes prescribed in specific standard test methods. In other cases, standards are established for given sectors of technology (for example, mechanical face seal technology). See the sections on specific wear types in this Handbook for more information.
- **wear rate (of seals)**
  - The amount of seal-surface wear, stated in terms of mils worn in some designated time period. *Note:* One commonly used unit is mils per hundred hours (Ref 4m).
- **wear resistance**
  - The resistance of a body to removal of material by wear processes, expressed as the reciprocal of wear rate. *Notes:* Wear resistance is a function of the conditions under which the wear process takes place. These conditions should always be carefully specified. Relative wear resistance is sometimes quoted, using arbitrary standards (Ref 1m).
- **wear scar**
  - The portion of a solid surface that exhibits evidence that material has been removed from it due to the influence of one or more wear processes.
- **wear transition**
  - Any change in the wear rate or in the dominant wear process occurring at a solid surface. *Note:* Wear transitions can be produced by an external change in the applied conditions (for example, load, velocity, temperature, or gaseous environment) or by time-dependent changes (aging) of the materials and restraining fixtures in the tribosystem.
- **wedge effect**
  - The establishment of a pressure wedge in a lubricant (Ref 1). See also wedge formation (2).
- **wedge formation**

- (1) In sliding metals, the formation of a wedge or wedges of plastically sheared metal in local regions of interaction between sliding surfaces. *Note:* This type of wedge is also known as a prow. It is similar to a built-up edge (Ref 1). (2) In hydrodynamic lubrication, the establishment of a pressure gradient in a fluid flowing into a converging channel. *Note:* This is also known as wedge effect (Ref 1).

- **weepage**
  - A minute amount of liquid leakage by a seal. *Note:* It has rather arbitrary limits, but is commonly considered to be a leakage rate of less than one drop of liquid per minute (Ref 4m).
- **welding**
  - In tribology, adhesion between solid surfaces in direct contact at any temperature (Ref 1).
- **whirl (oil)**
  - Instability of a rotating shaft associated with instability in the fluid film (Ref 1).
- **white layer**
  - In tribology, a white-etching layer, typically associated with ferrous alloys, that is visible in metallographic cross sections of bearing surfaces. *Note:* Some interpretations of white layers involve the presence of amorphous material; however, in other cases, the layer is thought to be finely microcrystalline such that no structure is visible by optical microscopy. See also Beilby layer and highly deformed layer .
- **white metal (whitemetal)**
  - A light-colored bearing alloy based on tin or lead. *Note:* Babbitt metal is one of the white metals (Ref 1).
- **wick lubrication**
  - A system in which the lubricant is delivered to the bearing surface by means of a wick (Ref 1).
- **wiper**
  - A pad of felt or other material used to supply lubricant or to remove debris (Ref 1).
- **wiping**
  - In tribology, the smearing or removal of material from one point, often followed by the redeposition of the material at another point, on the surface of two bodies in sliding contact. *Note:* The smeared metal is usually softened or melted (Ref 1).
- **work factor**
  - A measure of the stability of a lubricant when subjected to an endurance test. *Note:* The work factor is expressed as the average value of the ratio of three characteristics (viscosity, carbon residue, and neutralization number) as measured before the test to those same characteristics as measured after the test (Ref 5).
- **wrapped bush (bearing)**
  - A thin-walled steel bush lined with a bearing alloy, or any other bearing bush made from strip (Ref 1).
- **wrist pin bearing**
  - The bearing at the crankshaft end of an articulated connecting rod in a "V" engine (Ref 1).

References

1. *Glossary of Terms and Definitions in the Field of Friction, Wear and Lubrication (Tribology),* Research Group on Wear of Engineering Materials, Organisation for Economic Cooperation and Development (OECD), Paris, 1969
2. "Standard Terminology Relating to Erosion and Wear," G 40, *Annual Book of ASTM Standards,* ASTM
3. M.B. Peterson and W.O. Winer, Ed., *Wear Control Handbook,* American Society for Mechanical Engineers, 1980
4. *A Glossary of Seal Terms,* Special Publication SP-1, Society of Tribologists and Lubrication Engineers, 1983
5. E.L. McMurtrey, Ed., *High Performance Solid and Liquid Lubricants,* Noyes Data Corp., 1987, p 399-402
6. E.A. Avallone and T. Baumeister III, Ed., *Marks Standard Handbook for Mechanical Engineers,* 9th ed., 1987, p 8-133 to 8-135
7. E.E. Bisson and W.J. Anderson, "Advanced Bearing Technology," Special Publication SP-38, National Aeronautics and Space Administration, 1964

8. W.A. Glaeser, Battelle Columbus Laboratories, private communication, 1990
9. K.E. Speer, Diamond-Ceramic Coating of the Future, *J. Am. Ceram. Soc.,* Vol 72 (No. 2), 1989, p 171-191
10. D. Godfrey, *Lubr. Eng.,* Dec 1989, p 750-751
11. *Compilation of ASTM Standards,* 6th ed., ASTM, 1986
12. *Interpreting Service Damage in Rolling Type Bearings,* American Society of Lubrication Engineers (now Society of Tribologists and Lubrication Engineers), 1986
13. "Determination of Slurry Abrasivity (Miller Number) and Slurry Abrasion Response (SAR Number)," G 75, *Annual Book of ASTM Standards,* ASTM
14. A.D. Sarkar, *Friction and Wear,* Academic Press, London, 1980, p 342-343

Abbreviations, Symbols, and Tradenames

Abbreviations and Symbols

- ***a***
  - contact radius for circular contact; semimajor axis of elliptical contact area; area
- a_1
  - life adjustment factor for reliability; reliability factor for other than 90% reliability; reliability factor to adjust the L_{10} life
- a_2
  - life adjustment factor for special bearing properties; material factor for other than standard bearing steels
- a_3
  - application or environmental factor, primarily recognizing the lubricant condition within a bearing (can also include alignment, internal load distribution changes, temperature, contaminants [liquid and solid], and fatigue spall criteria); life adjustment factor for operating conditions
- **A**
  - austenite; ampere; air
- ***A***
  - cross-sectional area; contact area; projected area; solid material
- Å
  - angstrom
- A_1
  - cross-sectional area of a scratch
- **AA**
  - arithmetic roughness; atomic absorption; Aluminum Association
- **ABS**
  - acrylonitrile butadiene styrene
- A_c
  - contact area
- **ac**
  - alternating current
- **AC**
  - air cool
- Ac_1
  - temperature at which austenite begins to form during heating
- Ac_3
  - temperature at which transformation of ferrite to austenite is completed during heating

- **Ac_{cm}**
  - in hypereutectoid steel, temperature at which cementite completes solution in austenite
- **ACF**
  - autocorrelation function
- **A_D**
  - cross-sectional area of displaced material
- **A_e**
  - surface area
- **Ae_{cm},Ae_1, Ae_3**
  - equilibrium transformation temperatures in steel
- **AEF**
  - average energy flux
- **AEM**
  - analytical electron microscope/microscopy
- **AES**
  - Auger electron spectroscopy; acoustic emission spectroscopy
- **AFM**
  - atomic force microscope
- **AG**
  - aged; alternating grinding
- **AGMA**
  - American Gear Manufacturers Association
- **AGR**
  - advanced gas-cooled reactor
- **AISI**
  - American Iron and Steel Institute
- **A_{LB}**
  - load-bearing area (that is, vertical projection of the contact area)
- **alcop**
  - aluminum-copper bronze
- **Ams**
  - Amsler circumferential, rotating disk-on-disk machine
- **AMS**
  - Aerospace Material Specification; acoustic material signature
- **A_N**
  - nominal area of contact
- **AN**
  - annealed
- **ANSI**
  - American National Standards Institute
- **A_p**
  - cross-sectional area of a groove
- **API**
  - American Petroleum Institute
- **A_r**
  - real area of contact; cross-sectional area of formed ridges
- **AR**
  - abrasion resistant
- **Ar_1**
  - temperature at which transformation to ferrite or to ferrite plus cementite is completed on cooling
- **Ar_3**
  - temperature at which transformation of austenite to ferrite begins on cooling
- **Ar_{cm}**
  - temperature at which cementite begins to precipitate from austenite on cooling
- **ARE**
  - activated reactive evaporation
- **ARP**

    - Aerospace Recommended Practice
- **ASLE**
    - American Society of Lubrication Engineers (former name of STLE)
- **ASME**
    - American Society of Mechanical Engineers
- **ASTM**
    - American Society for Testing and Materials
- **at.%**
    - atomic percent
- **ATF**
    - automatic transmission fluid
- **atm**
    - atmospheres (pressure)
- A_w
    - cross-sectional area of the removed material
- **AWL**
    - Archard wear law
- **AWN**
    - antiwear number
- **AWS**
    - American Welding Society
- AX
    - gaseous compound
- b
    - journal bearing width; groove width; semi-minor axis of elliptical contact area
- b_1
    - thermal contact coefficient of body 1
- b_2
    - thermal contact coefficient of body 2
- **B**
    - bainite
- B
    - gaseous reaction product
- B_{10} **life**
    - bearing life
- **bal**
    - balance
- **BAM**
    - Bundesanstalt für Materialforschung und -Prufung
- **BARE**
    - biased activated reactive evaporation
- **bcc**
    - body-centered cubic
- **bct**
    - body-centered tetragonal
- **BEP**
    - best efficiency point (pump flow rate)
- b_f
    - thermal contact coefficient of film
- b_H
    - semiwidth of Hertzian contact band
- b_m
    - rating factor for contemporary, normally used material and manufacturing quality, the value of which varies with bearing type and design
- B_M
    - thermal contact coefficient
- **BOR**

- flat block pressed against the cylindrical surface of a rotating ring

- b_s
  - thermal contact coefficient of a substrate
- **BUE**
  - built-up edge
- c
  - edge length in crystal structure; speed of light; specific heat; constant; bearing radial clearance; conductivity
- **C**
  - cementite; coulomb; heat capacity
- C
  - constant; velocity of light; bearing capacity; shock wave velocity
- C_0
  - acoustic velocity
- C_a
  - basic radial load rating for thrust ball bearings
- **CAD/CAM**
  - computer-aided design/computer-aided manufacturing
- **CAE**
  - corrosion-affected erosion
- **CANMET**
  - Canada Center for Mineral and Energy Technology
- **CBED**
  - convergent-beam electron diffraction
- **CBN**
  - cubic boron nitride
- **CCT**
  - continuous-cooling transformation
- **CCVD**
  - conventional chemical vapor deposition
- c_d
  - diametrical clearance
- **cd**
  - candela
- **CE**
  - carbon equivalent
- C_f
  - bearing friction factor
- **CFG**
  - creep feed grinding
- **cgs**
  - centimeter-gram-second (system of units)
- **CI**
  - compression ignition
- **CLA**
  - center-line average
- c_M
  - specific heat per unit mass
- **cm**
  - centimeter
- **CMA**
  - cylindrical mirror analyzer
- **CNC**
  - computer numerical control
- C_{oa}
  - basic static axial load rating
- C_{or}

- basic static radial load rating
- **c_p**
  - specific heat at constant pressure
- **cP**
  - centipoise
- **CPI**
  - chemical process industry
- **c_r**
  - radial clearance
- **C_r**
  - basic radial load rating
- **C_R**
  - basic radial dynamic capacity based on the equations
- **CR**
  - cold reduced
- **CRT**
  - cathode-ray tube
- **CSOM**
  - confocal scanning optical microscope
- **cSt**
  - centistokes
- **C_T**
  - basic thrust dynamic capacity
- **CTE**
  - coefficient of thermal expansion
- **CTEM**
  - conventional transmission electron microscope/microscopy
- **CVD**
  - chemical vapor deposition
- **CVI**
  - chemical vapor infiltration
- **CVM**
  - consumable vacuum-melted
- **CVN**
  - Charpy V-notch (impact test or specimen)
- **cw**
  - continuous wave
- **d**
  - day
- ***d***
  - used in mathematical expressions involving a derivative (denotes rate of change); depth; diameter
- **D**
  - die cast
- ***D***
  - diameter; penetration depth
- ***d****
  - average value of two diagonals in Vickers indenter
- ***da/dN***
  - fatigue crack growth rate
- **DAT**
  - digital audio tape
- **d_B**
  - diameter of Brinell hardness indentation
- **dB**
  - decibel
- **D_B**
  - ball diameter of Brinell hardness tester

- **DBT**
  - ductile-to-brittle transition
- **dc**
  - direct current
- d_f
  - length of hardness indentation diagonal after wear occurs
- D_f
  - thermal diffusivity of lubricant film
- dF_N/dt
  - loading rate
- **dhcp**
  - double hexagonal close-packed
- D_i
  - thermal diffusivity of body i
- **diam**
  - diameter
- **DIN**
  - Deutsche Industrie-Normen (German Industrial Standards)
- **DLC**
  - diamondlike carbon
- dL/dX
  - loading rate
- $DN_e/A_e dt$
  - number of molecules evaporating from a surface area A_e in time dt
- d_o
  - diameter resulting from elastic deformation; width of diameter of area supporting a load before wear (that is, Hertz width); elastic contact diameter; length of the diagonal of a hardness indentation before wear occurs
- **dpa**
  - displacements per incident atom
- **DPH**
  - diamond pyramid hardness
- d_{v1}
  - one of two diagonals in Vickers indenter
- d_{v2}
  - one of two diagonals in Vickers indenter
- dx/dt
  - scratching speed
- e
  - natural log base, 2.71828; charge of an electron; specific grooving energy
- E
  - erosion rate; Young's modulus; modulus of elasticity; applied voltage; energy; activation energy
- E_1
  - modulus of elasticity of a pinion
- E_2
  - modulus of elasticity of a gear
- **EAP**
  - end of active profile
- **EAW**
  - electric arc wire (spray)
- e_B
  - maximum height of asperities on a bearing
- **EB**
  - electron beam
- E_c
  - Young's modulus of a coating
- **E/C**

- erosion/corrosion
- **ECP**
  - electron channeling pattern
- $\boldsymbol{E}_{\mathbf{d}}$
  - displacement energy
- **EDM**
  - electrical discharge machining
- **EDS**
  - energy-dispersive spectrometer
- **EDX**
  - energy-dispersive spectroscopy
- **EDXA**
  - energy-dispersive x-ray analysis
- **EEC**
  - erosion-enhanced corrosion
- **EELS**
  - electron energy loss spectroscopy
- **EHD**
  - elastohydrodynamic
- **EHL**
  - elastohydrodynamic lubrication
- $\boldsymbol{E}_{\mathbf{i}}$
  - kinetic energy of an incident ion
- ***EI***
  - flexural rigidity
- **ELI**
  - extra-low interstitial
- **ELNES**
  - energy-loss near-edge structure
- **emf**
  - electromotive force
- **EP**
  - extreme pressure
- **EPA**
  - Environmental Protection Agency
- **EPMA**
  - electron probe microanalysis
- **Eq**
  - equation
- $\boldsymbol{E}_{\mathbf{r}}$
  - reduced modulus of elasticity
- $\boldsymbol{e}_{\mathbf{s}}$
  - maximum height of asperities on a bearing shaft
- **ESCA**
  - electron spectroscopy for chemical analysis
- **ESD**
  - electrospark deposition; electron-stimulated desorption
- **ESDIED**
  - electron-stimulated desorption-ion energy distribution
- **ESR**
  - electroslag remelting
- ***et al.***
  - and others
- **eV**
  - electron volt
- **exp**
  - base of the natural logarithm

- ***f***
  - frequency; transfer; minimum facet contact area fraction; ratio of raceway curvature radius to ball diameter; friction coefficient; volume fraction
- $\boldsymbol{f_1}$
  - factor depending on bearing design and load
- **F**
  - ferrite
- ***F***
  - face width of line contact in direction perpendicular to rolling; bushing length; coefficient dependent on the form of the heat flux distribution; friction force; load; force
- $\mathbf{F_o}$
  - Fourier modulus
- $\boldsymbol{F_a}$
  - applied thrust load
- $\boldsymbol{F_A}$
  - adhesive friction force
- **FAA**
  - Federal Aviation Administration
- $\boldsymbol{f_{ab}}$
  - removal coefficient
- **FBC**
  - fluidized bed combustion
- $\boldsymbol{f_c}$
  - volume fraction of coating material being deformed; factor that depends on the composition of the material geometry of bearing components and the accuracy to which the various components are made
- **fcc**
  - face-centered cubic
- **fct**
  - face-centered tetragonal
- **FEA**
  - finite element analysis
- $\boldsymbol{F_F}$
  - frictional force
- **FFT**
  - fast Fourier transform
- **Fig**
  - figure
- $\boldsymbol{F_N}$
  - applied (normal) load
- **FOF**
  - flat surface sliding on another flat surface
- $\boldsymbol{F_P}$
  - plowing component of friction
- $\boldsymbol{F_r}$
  - applied radial load
- $\boldsymbol{f_s}$
  - volume fraction of substrate material being deformed
- **FSZ**
  - fully stabilized zirconia
- **ft**
  - foot
- $\boldsymbol{F_T}$
  - tangential force
- **FTIR**
  - Fourier transform infrared spectroscopy
- $\boldsymbol{f_v}$

    - constant dependent on bearing type and lubrication
- **FZG**
    - Forschungsstelle für Zahnräder und Getriebebau (Technical Institute for the Study of Gears and Drive Mechanisms)
- $F\beta$
    - applicable load
- **g**
    - gram
- ***g***
    - acceleration due to gravity
- **G**
    - graphite; gauss
- ***G***
    - dimensionless material parameter
- $G\chi$
    - limiting shear modulus
- **gal**
    - gallon
- **gf**
    - gram force
- G_{Ic}
    - critical strain energy release rate
- **GPa**
    - gigapascal
- **GTA**
    - gas tungsten arc
- **h**
    - hour; vertical thickness; minimum lubricant film thickness; bearing film thickness; depth of wear; Planck's constant. (6.626×10^{-27} erg · s)
- $\bar{h}$
    - average lubricant film thickness
- ***h****
    - asperity film thickness or micro-elastohydrodynamic film thickness
- h_0
    - inlet film thickness
- **H**
    - Henry
- ***H***
    - enthalpy; hardness; magnetic field; height; depth of wear; power loss
- H_a
    - allowable depth of wear
- **HAIM**
    - high-frequency acoustic imaging
- **HAZ**
    - heat-affected zone
- **HB**
    - Brinell hardness
- H_c
    - coating hardness; composite hardness
- **hcp**
    - hexagonal close-packed
- H_{comp}
    - effective hardness number
- H_{def}
    - hardness of highly deformed structure
- **HDPE**

    - high-density polyethylene
- **h_e**
    - depth of wear with elastic springback
- **HERF**
    - high-energy-rate forming
- **H_f**
    - hardness of film
- **HIP**
    - hot isostatic pressing
- **HK**
    - Knoop hardness
- **HLB**
    - hydrophile-lipophile balance
- **h_{min}**
    - minimum oil film thickness
- **HMX**
    - cyclotetramethylene tetranitramine
- **HOPG**
    - highly oriented pyrolytic graphite
- **hp**
    - horsepower
- **HP**
    - high-purity; hot pressed
- **H_P**
    - plowing stress
- **HPC**
    - high-pressure compressor
- **HPSTC**
    - highest point of single tooth contact
- **HPT**
    - high-pressure turbine
- **HR**
    - Rockwell hardness (Requires scale designation, such as HRC for Rockwell C hardness)
- **HREM**
    - high-resolution electron microscopy (microhardness)
- **HRMF**
    - Rockwell microficial (microhardness)
- **H_s**
    - substrate hardness
- **H_S**
    - scratch hardness
- **HSLA**
    - high-strength low-alloy
- **HSS**
    - high-speed steel(s)
- **H/t**
    - rate of dry wear
- **HTX**
    - high-temperature crystalline (polymer)
- **HV**
    - Vickers hardness
- **HVOF**
    - high-velocity oxyfuel (thermal-spray coating)
- **HV_P**
    - Vickers hardness at applied force of P
- **Hz**
    - hertz

- ***i***
  - number of rows of balls in a bearing
- ***I***
  - intensity; electrical current; bias current
- $\boldsymbol{I}_{\mathbf{bs}}$
  - backscattered current
- $\boldsymbol{I}_{\mathbf{c}}$
  - specimen current
- **ICPE**
  - inductively coupled plasma emission
- **ID**
  - inside diameter
- $\boldsymbol{I}_{\mathbf{diff}}$
  - diffracted-beam current
- $\boldsymbol{I}_{\mathbf{i}}$
  - incident-beam current
- **I/M**
  - ingot metallurgy
- **in.**
  - inch
- **IP**
  - conventional ion plating
- **IPN**
  - interpenetrating polymer network
- **IPTS**
  - International Practical Temperature Scale
- **IR**
  - infrared
- **IRG**
  - International Research Group
- $\boldsymbol{I}_{\mathbf{s}}$
  - emitted secondary electron current
- **IS**
  - sliding down an inclined surface
- **ISCC**
  - intergranular stress-corrosion cracking
- **ISO**
  - International Organization for Standardization
- **ISS**
  - ion scattering spectroscopy
- $\boldsymbol{I}_{\mathbf{t}}$
  - transmitted electron current
- **J**
  - joule
- ***J***
  - ion current density
- $\boldsymbol{J}_{\mathbf{eq}}$
  - Jominy equivalent
- **k**
  - Boltzmann constant
- ***k***
  - thermal conductivity; Boltzmann constant; wear factor; wear coefficient
- **K**
  - Kelvin
- ***K***
  - wear coefficient; specific wear rate
- **kg**

- kilogram
- **kgf**
  - kilogram force
- K_{Ic}
  - plane-strain fracture toughness
- K_{ISCC}
  - threshold stress intensity to produce stress-corrosion cracking
- **km**
  - kilometer
- **kN**
  - kilonewton
- **kPa**
  - kilopascal
- k_s
  - specific cutting force
- **ksi**
  - kips (1000 lbf) per square inch
- **kV**
  - kilovolt
- **kW**
  - kilowatt
- l
  - length
- ℓ
  - length
- ℓ_0
  - length of asperity
- **L**
  - longitudinal, liter
- L
  - relative bearing life; length of bearing (land) of a die; contact length
- L_{10}
  - fatigue life that 90% of a group of bearings will endure; bearing life for 90% reliability
- L_{50}
  - median bearing life
- **lb**
  - pound
- **lbf**
  - pound force
- L_c
  - characteristic length
- **LCVD**
  - laser-induced chemical vapor deposition
- **LCL**
  - lower control limit
- **LCP**
  - liquid crystal polymer
- **LCSM**
  - laser confocal scanning microscope
- **LDPE**
  - low-density polyethylene
- L_{min}
  - minimum contact length
- **In**
  - natural logarithm (base *e*)
- L_{na}

- calculated life adjusted by the multiplicative factors a_1, a_2, and a_3 (in rev); the bearing life in millions of revolutions (na × 10^6

- **log**
  - common logarithm (base 10)
- **LPCVD**
  - low-pressure chemical vapor deposition
- **LPG**
  - liquified petroleum gas
- **LPSTC**
  - lowest point of single tooth contact
- **LPT**
  - low-pressure turbine
- **L_R**
  - bearing life in terms of ratio to catalog expected life (that is, relative fatigue life)
- **LSI**
  - large-scale integrated (circuit)
- **LT**
  - long transverse (direction)
- **LVDT**
  - linear variable differential transformer
- **m**
  - meter
- ***m***
  - constant shear, or frictional factor, Weibull slope, for reliability above 90%; Weibull distribution; dispersion parameter; molecular weight
- ***M***
  - bearing total friction torque
- **M_1**
  - bearing friction torque due to rolling element-raceway contact loading
- **mA**
  - milliampere
- **M_A**
  - atomic weight of target material
- **MAS**
  - Microbeam Analysis Society
- **MCD**
  - magnetic chip detection
- **MEP**
  - mean effective pressure
- **MeV**
  - megaelectronvolt
- **M_f**
  - temperature at which martensite formation finishes during cooling
- **MFP**
  - mean free path
- **mg**
  - milligram
- **Mg**
  - megagram (metric tonne, or kg × 10_3)
- **min**
  - minute; minimum
- **M_i**
  - mass of the incident atom
- **MIG**
  - metal inert gas (welding)
- **MIR**
  - multiple internal reflectance

- **mL**
  - milliliter
- **mm**
  - millimeter
- **MMC**
  - metal-matrix composite
- **MOCVD**
  - metal-organic chemical vapor deposition
- **mPa**
  - millipascal
- **MPa**
  - megapascal
- **mpg**
  - miles per gallon
- **mph**
  - miles per hour
- **MRR**
  - material removal rate
- **ms**
  - millisecond
- **M_s**
  - temperature at which martensite starts to form from austenite on cooling
- **MS**
  - megasiemens; magnetron sputtering
- **MSA**
  - Microscopy Society of America
- **MSDS**
  - material safety data sheet
- **mT**
  - millitesla
- **M_t**
  - mass of the target atom
- **MTTF**
  - mean time to failure
- **mV**
  - millivolt
- **MV**
  - megavolt
- **M_v**
  - bearing friction torque due to hydrodynamic fluid friction
- ***n***
  - pinion speed; load life exponent (experimentally based, with consensus values published in the bearing standards; typically, n = 3 for ball bearings and n = 10/3 for roller bearings); number of triangles in regular polygon; independent contact points conducting in parallel; bearing speed
- **N**
  - newton
- ***N***
  - number of cycles; normal solution; angular velocity of cylindrical contact; bearing speed; normal force
- **NA**
  - numerical aperture
- **NASA**
  - National Aeronautics and Space Administration
- **NBS**
  - National Bureau of Standards (former name of NIST)
- **NDE**
  - nondestructive evaluation

- **NER**
  - erosion resistance number
- n_i
  - inner ring speed
- **NIST**
  - National Institute of Standards and Technology
- **nm**
  - nanometer
- n_m
  - cage speed (rolling-element orbital speed)
- **NMMA**
  - National Marine Manufacturers Association
- n_o
  - outer ring speed
- **No.**
  - number
- N_0
  - rationalized incubation period
- ***NOR***
  - incubation resistance number
- **NPSH**
  - net positive suction head
- **NPSHA**
  - available net positive suction head
- **NPSHR**
  - required net positive suction head
- n_{RE}
  - ball or roller speed about its own axis
- **ns**
  - nanosecond
- **NSp**
  - not specified
- $N(\Lambda)/N_{cat}$
  - relative life factor
- $N_{\mu e=0}$
  - fatigue life when surface traction equals zero
- **Oe**
  - oersted
- **OECD**
  - Organisation for Economic Cooperation and Development
- **OFD**
  - oxyfuel detonation (spray)
- **OFP**
  - oxyfuel powder (spray)
- **OFW**
  - oxyfuel wire (spray)
- **OMCVD**
  - organo-metallic chemical vapor deposition
- **ORNL**
  - Oak Ridge National Laboratory
- **OSHA**
  - Occupational Safety and Health Administration
- **oz**
  - ounce
- **p**
  - page

- ***p***
  - pressure; hydrostatic pressure acting on the surface
- ***p****
  - local asperity contact pressure; equilibrium vapor pressure at an evaporant surface
- $\bar{p}$
  - average (bulk) hydrodynamic pressure
- **P**
  - pearlite
- ***P***
  - specific load or unit load; pressure; transmitted power
- P_a
  - absolute ambient pressure
- $\bar{p}_a$
  - average (bulk) asperity contact pressure
- **Pa**
  - pascal
- **PA**
  - plasma arc (spray); prealloyed; polyamide
- **PACVD**
  - plasma-assisted chemical vapor deposition
- **PAN**
  - polyacrylonitrile
- **PAO**
  - polyalphaolefin
- **PAPVD**
  - plasma-assisted physical vapor deposition
- **PBT**
  - polybutylene terephthalate
- **PCD**
  - polycrystalline diamond
- **PCV**
  - positive crankcase ventilator
- **PDF**
  - probability density function
- **Pe**
  - Péclet number
- **PEEK**
  - polyetheretherketone
- **PEI**
  - polyetherimide
- **PEK**
  - polyetherketone
- **PEP**
  - passive extreme pressure
- **PES**
  - polyether sulfone
- **PETN**
  - pentaerithritol tetranitrate
- **PETP**
  - polyethylene terephthalate
- **PFPE**
  - polyperfluoroalkylether
- **pH**
  - negative logarithm of hydrogen-ion activity
- p_H
  - maximum Hertzian contact pressure
- **PH**

- precipitation hardenable
- **P_H**
  - hardness; Brinell pressure
- **PHL**
  - plastohydrodynamic lubrication
- **p_i**
  - pocket pressure in hydrostatic bearing
- **PKA**
  - primary knock-on atom
- **PLP**
  - percent of large particles
- **p_m**
  - flow pressure or hardness of material
- **PM**
  - permanent mold
- **P/M**
  - powder metallurgy
- **PMMA**
  - polymethyl methacrylate
- **P_N**
  - nominal normal stress on contact patch
- **p_o**
  - yield pressure
- **POD**
  - pin on disk
- **POF**
  - pin on flat
- **POM**
  - polyoxymethylene
- **P_{or}**
  - static equivalent radial load
- **POR**
  - pin sliding against the cylindrical surface of a rotating ring
- **ppb**
  - parts per billion
- **ppba**
  - parts per billion atomic
- **ppm**
  - parts per million
- **ppmm**
  - parts per million by mass
- **PPS**
  - polyphenylene sulfide
- **ppt**
  - parts per trillion
- **PSD**
  - power spectral density
- **psi**
  - pounds per square inch
- **psia**
  - pounds per square inch absolute
- **psig**
  - gage pressure (pressure relative to ambient pressure) in pounds per square inch
- **PSII**
  - plasma-source ion implantation
- **PSZ**
  - partially stabilized zirconia

- **PTA**
  - plasma transferred arc
- **PTFE**
  - polytetrafluoroethylene
- P_u
  - fatigue load limit
- **PVC**
  - polyvinyl chloride
- **PVD**
  - physical vapor deposition
- **PVDF**
  - polyvinylidene-difluoride
- q
  - heat flux distribution; oil flow rate
- Q
  - thermal energy generated per unit time
- q_{av}
  - average heat flux distribution
- $\overline{Q}_e$
  - contact stress
- Q_{gen}
  - heat generation
- Q_i
  - rate of heat supplied to body i
- r
  - radius; radial distance of receiver from source; resistivity
- **R**
  - roentgen
- R
  - radius; gas constant; reliability expressed in terms of percent survival; resistance
- **R**
  - force vector
- r_o
  - relative radius at an area before wear
- R_o
  - surface radius with lubricant film
- r_1
  - radius of surface 1 at area before wear
- r_2
  - radius of surface 2 at area before wear
- r_I
  - radius of rolling body I
- r_{II}
  - radius of rolling body II
- R_a
  - surface roughness in terms of arithmetic average
- ***RA***
  - reduction in area
- r^{β}
  - bushing radius
- **RB**
  - reaction bonded
- **RCF**
  - rolling contact fatigue
- **RCW**
  - rolling contact wear

- **RDX**
  - cyclotrimethylene trinitramine
- $\boldsymbol{R}_{\mathbf{e}}$
  - equivalent radius of curvature; rationed erosion rate
- **RE**
  - rare earth
- **Ref**
  - reference
- **REF**
  - relative erosion factor
- **rf**
  - radio frequency
- **RH**
  - relative humidity
- **RIP**
  - reactive ion plating
- **rms**
  - root mean square
- $\boldsymbol{R}_{\mathbf{n}}$
  - neutral radius
- $\boldsymbol{R}_{\mathbf{p}}$
  - single predominant peak height; leveling depth
- **rpm**
  - revolutions per minute
- $\boldsymbol{R}_{\mathbf{pm}}$
  - mean height of highest peaks on five adjacent sampling lengths; average leveling depth
- **RPOF**
  - reciprocating pin on flat
- $\boldsymbol{R}_{\mathbf{q}}$
  - rms (root means square) roughness
- **R & O**
  - rust and oxidation inhibited
- $\boldsymbol{r}_{\mathbf{s}}$
  - shaft radius
- **RS**
  - reactive sputtering
- $\boldsymbol{R}_{\mathbf{sk}}$
  - skew roughness
- **RSOF**
  - reciprocating, spherically ended pin on a flat surface
- **RT**
  - room temperature
- **RTD**
  - resistance temperature detector
- **RTO**
  - rejected take-off
- **RTV**
  - room-temperature vulcanizing
- **RWAT**
  - rubber wheel abrasion test
- $\boldsymbol{R}_{\mathbf{y}}$
  - maximum peak-to-valley height
- $\boldsymbol{R}_{\mathbf{z}}$
  - ten-point height
- **s**
  - second
- $\boldsymbol{s}$

    - standard deviation
- **S**
    - sand cast; siemens
- ***S***
    - sputtering yield; Sommerfeld number; rotational speed; rms average surface roughness; probability of survival
- **SAD**
    - selected-area diffraction
- **SAE**
    - Society of Automotive Engineers
- **SAG**
    - semiautogenous
- **SAM**
    - scanning acoustic microscope/microscopy
- **SAP**
    - sintered aluminum powder; start of active profile
- **SAR**
    - slurry abrasion response (number)
- **SAW**
    - surface acoustic wave
- **SCaM**
    - scanning capacitance microscope
- **SCC**
    - stress-corrosion cracking
- **SEM**
    - scanning electron microscope/microscopy
- **SF**
    - segregation factor
- **SFA**
    - surface force apparatus
- **sfm**
    - surface feet per minute
- **SHC**
    - synthetic hydrocarbon
- **SHE**
    - standard hydrogen electrode
- **SI**
    - spark ignition; Système International d'Unités
- **SIMS**
    - secondary ion mass spectroscopy
- **Si/SiC**
    - silicon-infiltrated silicon carbide
- **SKA**
    - secondary knock-on atom
- **SLA**
    - surface layer activation
- **SLAM**
    - scanning laser acoustic microscope/microscopy
- **SMA**
    - shielded metal arc
- **SMAW**
    - shielded metal arc welding
- **SMD**
    - Sauter mean diameter
- **SNR**
    - signal-to-noise ratio
- **SOAP**

- spectrometric oil analysis program

- **SPC**
  - statistical process control
- **SPE**
  - solid particle erosion
- **SPOD**
  - spherically ended pin on flat disk
- **SPOF**
  - spherically ended pin on flat coupon
- **SPP**
  - second-phase particle
- **SPVF**
  - second-phase (particle) volume fraction
- **SQC**
  - statistical quality control
- **SRF**
  - strength reduction factor
- **SRM**
  - Standard Reference Materials
- **SRP**
  - self-reinforcing polymer
- **ST**
  - short transverse (direction)
- **STA**
  - solution-treated and aged
- **std**
  - standard
- **STEM**
  - scanning transmission electron microscope/microscopy
- **STLE**
  - Society of Tribologists and Lubrication Engineers (formerly ASLE)
- **STM**
  - scanning tunneling microscope/microscopy
- **StOD**
  - strand wrapped over a drum
- S_u
  - ultimate strength
- **SUS**
  - Saybolt Universal Seconds (viscosity)
- **Sv**
  - sievert
- t
  - thickness; time
- **T**
  - tesla
- T
  - temperature; thickness
- **TAN**
  - total acid number
- T_b
  - bulk volumetric temperature; gear bulk temperature (equilibrium temperature of surface of the gear teeth before they enter the meshing zone)
- T_{bi}
  - maximum bulk temperature
- **TBN**
  - total base number
- T_{btest}

- bulk temperature of test gears
- T_c
  - contact temperature; critical temperature; Curie temperature
- $\overline{T}_c$
  - average surface temperature over a given real area of contact
- **TC**
  - test configuration
- **TD**
  - Toyota Diffusion (process)
- T_e
  - electron temperature
- T_E
  - temperature rise of a bearing environment
- **TEI**
  - thermoelastic instability
- **TEM**
  - transmission electron microscope/microscopy
- T_f
  - flash temperature (local and instantaneous temperature rise that occurs on gear teeth as they pass through the meshing zone)
- $\overline{T}_f$
  - average flash temperature
- T_F
  - frictional torque
- T_{fi}
  - maximum flash temperature
- **TFTC**
  - thin-film thermocouple
- T_{ftest}
  - maximum flash temperature of test gears
- T_{fw}
  - surface temperature rise with film
- T_{fwo}
  - surface temperature rise without film
- **TGA**
  - thermogravimetric analysis
- **TGS**
  - threshold galling stress
- **TIG**
  - tungsten inert gas (welding)
- **TLA**
  - thin layer activation
- T_m
  - mean contact temperature; melting temperature
- **TPOD**
  - triple pin-on-disk
- t_r
  - required life
- **TRS**
  - transverse rupture strength
- T_s
  - temperature rise of a bearing surface; scuffing temperature
- **TS**
  - tribosystem
- $\overline{T}_s$
  - average (bulk) surface temperature rise

- $\boldsymbol{T^*_s}$
  - local asperity contact temperature rise
- **tsi**
  - tons per square inch
- **TSM**
  - tandem scanning microscope/microscopy
- **TTT**
  - time-temperature transformation
- **TW**
  - thrust washer test
- $\boldsymbol{u_I}$
  - surface velocity of body I in the rolling direction
- $\boldsymbol{u_{II}}$
  - surface velocity of body II in the rolling direction
- $\boldsymbol{U}$
  - heat of sublimation of target material; dimensionless speed parameter
- $\overline{U}$
  - rolling velocity
- **UCL**
  - upper control limit
- **UHDPE**
  - ultrahigh-density polyethylene
- **UHMWPE**
  - ultrahigh-molecular-weight polyethylene
- **UnSp**
  - unspecified method
- **UHV**
  - ultrahigh vacuum
- **UNS**
  - Unified Numbering System
- **USPHS**
  - United States Public Health Service
- **UTS**
  - ultimate tensile strength
- $\boldsymbol{v}$
  - velocity; particle velocity
- **V**
  - volt
- $\boldsymbol{V}$
  - sliding speed; velocity; volume
- $\boldsymbol{V_1}$
  - velocity of surface 1, which is tangential to contact and normal to contact length
- $\boldsymbol{V_2}$
  - velocity of surface 2, which is tangential to contact and normal to contact length
- **VAMAS**
  - Versailles Advanced Materials and Standards
- **VAR**
  - vacuum arc remelted
- **VCR**
  - video cassette recorder
- $\boldsymbol{V_e}$
  - worn volume with elastic springback; entraining velocity
- **VFD**
  - variable-frequency drive
- **VHN**
  - Vickers microindenter hardness number
- $\boldsymbol{V_i}$

- velocity of body i
- **VI**
  - viscosity index
- **VII**
  - viscosity index improver
- **VIM-VAR**
  - vacuum induction melted-vacuum arc remelted
- **vol**
  - volume
- **vol%**
  - volume percent
- v_{r1}
  - rolling velocity of pinion
- v_{r2}
  - rolling velocity of gear
- v_R
  - Rayleigh velocity
- **VTP**
  - vertical turbine pump
- **VTR**
  - video tape recorder
- w
  - width; mass losses
- **W**
  - watt
- W
  - load; bearing load-carrying parameter; weight of body; abrasive wear resistance; wear volume
- W_{12}
  - work of adhesion
- **WDS**
  - wavelength-dispersive spectrometer
- W_f
  - friction power loss
- $\overline{W}_{Nr}$
  - normal operating load
- **WPC**
  - wear particle concentration
- **WRA**
  - silicon carbide whisker-reinforced alumina
- **WRZTA**
  - silicon carbide whisker-reinforced, zirconia-toughened alumina
- W_s
  - shear power of deformation
- **wt%**
  - weight percent
- X
  - water concentration; interfacial constraint parameter
- **XPS**
  - x-ray photoelectron spectroscopy
- X_Γ
  - load sharing factor
- **XRD**
  - x-ray diffraction
- **XRF**
  - x-ray fluorescence
- **XTEM**
  - cross-sectional transmission electron microscopy

- **X_w**
    - welding factor
- $\bar{y}$
    - mean deviation
- ***Y***
    - yield stress in tension
- **yr**
    - year
- ***z***
    - incremental wear depth
- **z_1**
    - penetration depth
- ***Z***
    - number of rollers carrying load in one direction; number of balls carrying load in one direction; number of balls per row; atomic number
- **ZAP**
    - zone axis pattern
- **ZDDP**
    - zinc dialkyldithiophosphate
- **ZTA**
    - zirconia-toughened alumina
- °
    - angular measure; degree
- **°C**
    - degree Celsius (centigrade)
- **°F**
    - degree Fahrenheit
- $\rightleftarrows$
    - direction of reaction
- ÷
    - divided by
- =
    - equals
- $\approxeq$
    - approximately equals
- $\neq$
    - not equal to
- $\equiv$
    - identical with
- >
    - greater than
- $\gg$
    - much greater than
- $\geq$
    - greater than or equal to
- ∞
    - infinity
- $\propto$
    - is proportional to; varies as
- $\int$
    - integral of
- <
    - less than
- $\ll$
    - much less than
- $\leq$
    - less than or equal to
- ±

- maximum deviation
- -
  - minus; negative ion charge
- ×
  - diameters (magnification); multiplied by
- ·
  - multiplied by
- /
  - per
- %
  - percent
- +
  - plus; positive ion charge
- $\surd$
  - square root of
- $\sim$
  - approximately; similar to
- ∂
  - partial derivative
- α
  - nominal contact angle; semicone angle; thermal diffusivity; taper angle; angle of incidence; radius of clustered contact spots; contact patch radius, pressure-viscosity coefficient
- $\alpha\nu$
  - evaporation coefficient
- β
  - coefficient of friction
- γ
  - surface energy; pressure-viscosity coefficient; heat partition factor
- γ
  - shear strain rate
- γ_1
  - specific surface energy of coating
- γ_2
  - specific surface energy of substrate
- γ_{12}
  - interfacial specific free energy
- Γ^{-1}
  - composite of principal relative curvatures of contacting bodies
- δ
  - film width
- Δ
  - change in quantity; an increment; a range; critical amplitude
- ΔE
  - dissociation energy of gas molecules
- ϵ
  - lubricant film thickness; tribocontact parameter; emissivity; bearing eccentricity ratio; strain
- $\dot{\epsilon}$
  - strain rate
- η
  - lubricant viscosity; efficiency of material removal; degree of wear; abrasive fraction
- η_0
  - absolute viscosity at contact inlet; ambient viscosity of lubricant
- θ
  - angle
- λ
  - wavelength; pressure coefficient of viscosity; film thickness-to-roughness ratio; lambda ratio; specific film thickness
- λ_M

- heat conductivity
- Λ
  - specific film thickness; compressibility number; film thickness parameter
- μ
  - friction coefficient; magnetic permeability; lubricant viscosity at bearing operating temperature; absolute viscosity
- μ_0
  - absolute viscosity
- μ_A
  - adhesional component of the friction coefficient
- μ_e
  - surface traction
- μ_m
  - mean coefficient of friction
- μ_s
  - static coefficient of friction
- μ**F**
  - microfarads
- μ**in.**
  - microinch
- μ**m**
  - micrometer (micron)
- μ**s**
  - microsecond
- ν
  - Poisson's ratio; kinematic viscosity of a lubricant (in cSt) at bearing operating temperature
- ν_c
  - Poisson's ratio of coating material
- ν_1
  - Poisson's ratio of pinion
- ν_2
  - Poisson's ratio of gear
- ν_{40}
  - kinematic viscosity at 40 °C (105 °F)
- ξ
  - Poisson's ratio
- π
  - pi (3.141592)
- ρ
  - density
- ρ_M
  - density
- ρ_n
  - normal relative radius of curvature
- ρ_1
  - transverse radius of curvature of pinion
- ρ_2
  - transverse radius of curvature of gear
- σ
  - workpiece flow strength; stress; standard deviation; root-mean-square roughness; Stefan-Boltzmann constant
- σ_0
  - flow strength
- σ_1
  - rms surface roughness of pinion
- σ_2
  - rms surface roughness of gear

- σ_{max}
  - maximum normal contact stress at center of contact
- σ_s
  - shear stress
- σ_{xb}
  - extrusion pressure (back tension)
- σ_{xf}
  - drawing stress (front tension)
- Σ
  - summation of
- τ
  - applied stress
- $\overline{\tau}$
  - average lubricant shear stress
- τ*
  - local asperity contact shear stress
- τ_L
  - limiting shear stress
- Υ
  - rate of wear
- ϕ
  - standard normal distribution
- Φ
  - energy rate; work function
- Φ_0
  - angular contact width
- Ψ
  - rolling angle
- Ψ_b
  - base helix angle
- ω
  - angular velocity; frequency
- ω_1
  - angular velocity of pinion
- ω_2
  - angular velocity of gear
- ω_{Nr}
  - normal unit load
- Ω
  - ohm

o **Greek Alphabet**

- **A,** α
  - alpha
- **B,** β
  - beta
- Γ, γ
  - gamma
- Δ, δ
  - delta
- **E,** ϵ
  - epsilon
- **Z,** ζ
  - zeta
- **H,** η
  - eta
- Θ, θ

    - theta
- **I,** ι
    - iota
- **K,** κ
    - kappa
- Λ, λ
    - lambda
- **M,** μ
    - mu
- **N,** ν
    - nu
- Ξ, ξ
    - xi
- **O, o**
    - omicron
- Π, π
    - pi
- **P,** ρ
    - rho
- Σ, σ
    - sigma
- **T,** τ
    - tau
- Υ, υ
    - upsilon
- Φ, ϕ
    - phi
- **X,** χ
    - chi
- Ψ, ψ
    - psi
- Ω, ω,
    - omega

- **Tradenames**
- **20CB-3**
    - is a registered tradename of Carpenter Technology Corporation
- **Alloys 6B, 6K, 25 (L-605), 188, and Ultimet alloy 1233**
    - are registered tradenames of Haynes International, Inc.
- **Astralloy-V**
    - is a registered tradename of Plant City Steel Company
- **CBS-600**
    - is a registered tradename of The Timken Company
- **Endogas**
    - is a registered tradename of Seco/Warwick Corporation
- **Endox 214**
    - is a registered tradename of Enthrone, Inc.
- **Hastelloy C-276**
    - is a registered tradename of Enpar Sonderwerkstoffe GmbH
- **Inconel**
    - is a product of Inco Alloys International, Inc.
- **IRECA**
    - is a tradename of Institute de Recherche de Hydro Quebec
- **Lexan**
    - is a tradename of the General Electric Company
- **Magnagold**

    - is a product of General Magnaplate, Inc.
- **MetalLife**
    - is a registered tradename of Badger Metal Tech., Inc.
- **MP35N and MP159**
    - are registered tradenames of SPS Technologies Inc.
- **Nitronic**
    - is a registered tradename of Armco Steel Corporation
- **Nuclepore**
    - is a registered tradename of Nuclepore Corporation
- **Teflon**
    - is a registered tradename of E.I. DuPont de Nemours & Company, Inc.
- **Xydar**
    - is a registered tradename of Amoco Performance Products, Inc.